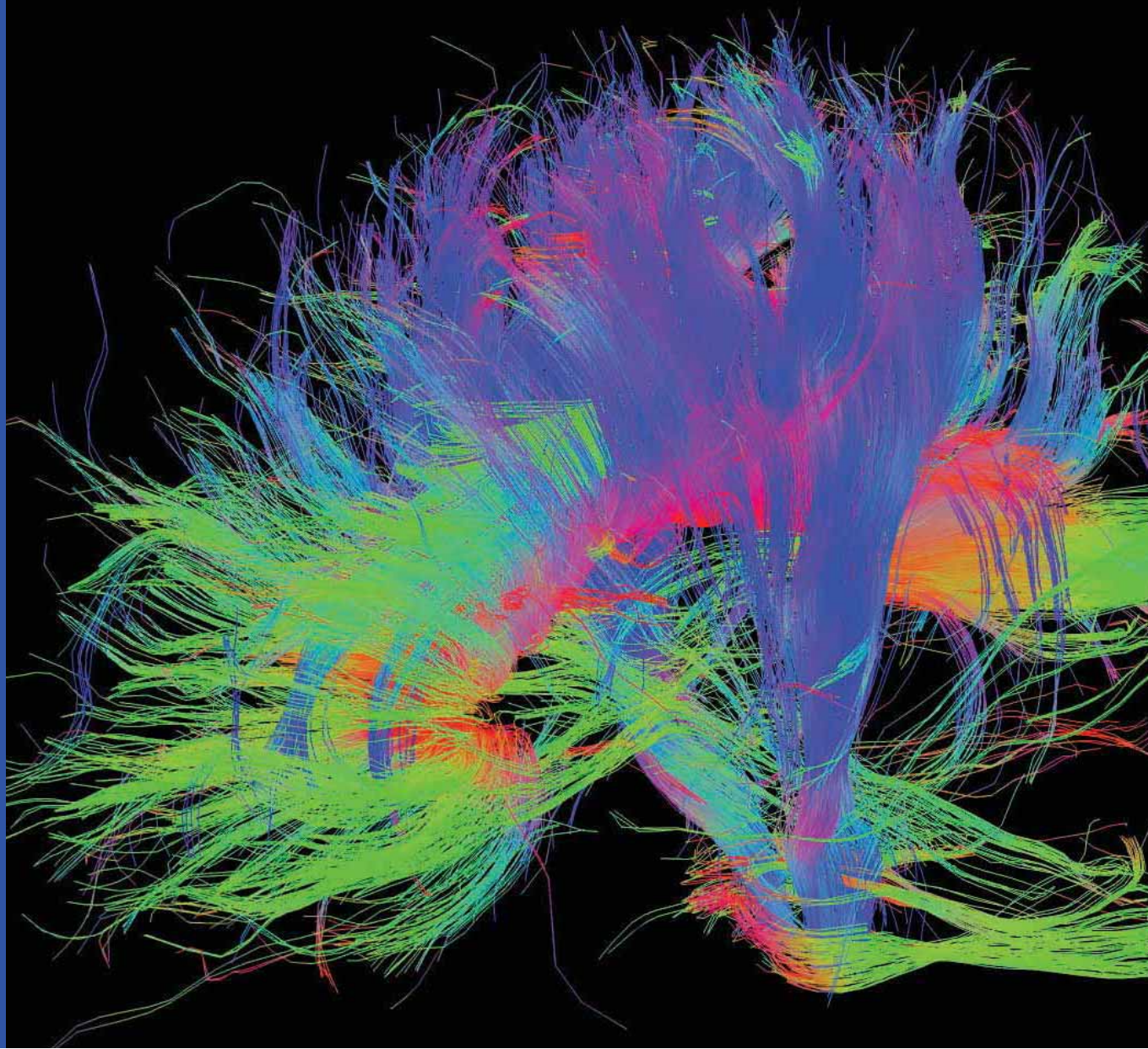
MICHAEL S. GAZZANIGA, EDITOR-IN-CHIEF

FOURTH EDITION

The Cognitive Neurosciences



THE COGNITIVE NEUROSCIENCES

THE COGNITIVE NEUROSCIENCES

Fourth Edition

Michael S. Gazzaniga, *Editor-in-Chief*

Section Editors:
Emilio Bizzi
Alfonso Caramazza
Leo M. Chalupa
Scott T. Grafton
Todd F. Heatherton
Christof Koch
Joseph E. LeDoux
Steven J. Luck
George R. Mangun
J. Anthony Movshon
Helen Neville
Elizabeth A. Phelps
Pasko Rakic
Daniel L. Schacter
Mriganka Sur
Brian A. Wandell

A BRADFORD BOOK
THE MIT PRESS
CAMBRIDGE, MASSACHUSETTS
LONDON, ENGLAND

© 2009 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

For information about special quantity discounts, please email special_sales@mitpress.edu

This book was set in Baskerville by SNP Best-set Typesetter Ltd., Hong Kong.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

The cognitive neurosciences/edited by Michael S. Gazzaniga. —4th ed.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-262-01341-3 (hardcover : alk. paper)
1. Cognitive neuroscience. I. Gazzaniga, Michael S.
[DNLM: 1. Brain—physiology. 2. Mental Processes—physiology. 3. Neurosciences. WL 300 C6766 2009]
QP360.5.N4986 2009
612.8′2—dc22 2009000145

10 9 8 7 6 5 4 3 2 1

For Charlotte Smylie Gazzaniga with deep appreciation and gratitude

CONTENTS

Preface xv

I DEVELOPMENT AND EVOLUTION

Introduction Pasko Rakic and Leo M. Chalupa 3

1 Development of the Primate Cerebral Cortex Pasko Rakic, Jon I. Arellano, and Joshua Breunig 7

2 Early Development of Neuronal Circuitry of the Human Prefrontal Cortex Ivica Kostović and Miloš Judaš 29

3 The Cognitive Neuroscience of Human Uniqueness Todd M. Preuss 49

4 Unraveling the Role of Neuronal Activity in the Formation of Eye-Specific Connections Leo M. Chalupa and Andrew D. Huberman 67

5 Brain Changes Underlying the Development of Cognitive Control and Reasoning Silvia A. Bunge, Allyson P. Mackey, and Kirstie J. Whitaker 73

II PLASTICITY

Introduction Helen Neville and Mriganka Sur 89

6 Patterning and Plasticity of Maps in the Mammalian Visual Pathway Sam Horng and Mriganka Sur 91

7 Synaptic Plasticity and Spatial Representations in the Hippocampus
Jonathan R. Whitlock and Edvard I. Moser 109

8 Visual Cortical Plasticity and Perceptual Learning Wu Li and
Charles D. Gilbert 129

9 Characterizing and Modulating Neuroplasticity of the Adult Human Brain
Alvaro Pascual-Leone 141

10 Exercising Your Brain: Training-Related Brain Plasticity Daphne Bavelier,
C. Shawn Green, and Matthew W. G. Dye 153

11 Profiles of Development and Plasticity in Human Neurocognition
Courtney Stevens and Helen Neville 165

III ATTENTION

Introduction Steven J. Luck and George R. Mangun 185

12 Attention: Theoretical and Psychological Perspectives Anne Treisman 189

13 Mechanisms of Selective Attention in the Human Visual System:
Evidence from Neuroimaging Sabine Kastner, Stephanie A. McMains,
and Diane M. Beck 205

14 The Frontoparietal Attention Network Maurizio Corbetta,
Chad M. Sylvester, and Gordon L. Shulman 219

15 Spatiotemporal Analysis of Visual Attention Jens-Max Hopf,
Hans-Jochen Heinze, Mircea A. Schoenfeld, and Steven A. Hillyard 235

16 Integration of Conflict Detection and Attentional Control Mechanisms:
Combined ERP and fMRI Studies George R. Mangun, Clifford D. Saron,
and Bong J. Walsh 251

17 A Right Perisylvian Neural Network for Human Spatial Orienting
Hans-Otto Karnath 259

18 Spatial Deficits and Selective Attention Lynn C. Robertson 269

19 The Effect of Attention on the Responses of Individual Visual Neurons
John H. R. Maunsell 281

20 Selective Attention Through Selective Neuronal Synchronization
Thilo Womelsdorf and Pascal Fries 289

IV SENSATION AND PERCEPTION

Introduction J. Anthony Movshon and Brian A. Wandell 305

21 Grandmother Cells, Symmetry, and Invariance: How the Term Arose and What the Facts Suggest Horace Barlow 309

22 Olfaction: From Percept to Molecule Yaara Yeshurun, Hadas Lapid, Rafi Haddad, Shani Gelstien, Anat Arzi, Lee Sela, Aharon Weisbrod, Rehan Khan, and Noam Sobel 321

23 Auditory Masking with Complex Stimuli Virginia M. Richards and Gerald Kidd, Jr. 343

24 Insights into Human Auditory Processing Gained from Perceptual Learning Beverly A. Wright and Yuxuan Zhang 353

25 Auditory Object Analysis Timothy D. Griffiths, Sukhbinder Kumar, Katharina von Kriegstein, Tobias Overath, Klaas E. Stephan, and Karl J. Friston 367

26 The Cone Photoreceptor Mosaic in Normal and Defective Color Vision Joseph Carroll, Geunyoung Yoon, and David R. Williams 383

27 Bayesian Approaches to Color Vision David H. Brainard 395

28 Wiring of Receptive Fields and Functional Maps in Primary Visual Cortex Dario L. Ringach 409

29 Encoding and Decoding with Neural Populations in the Primate Cortex Eyal Seidemann, Yuzhi Chen, and Wilson S. Geisler 419

30 Perceptual Filling-in: From Experimental Data to Neural Network Modeling Rainer Goebel and Peter De Weerd 435

31 Neural Transformation of Object Information by Ventral Pathway Visual Cortex Charles E. Connor, Anitha Pasupathy, Scott Brincat, and Yukako Yamane 455

32 The Cognitive and Neural Development of Face Recognition in Humans Elinor McKone, Kate Crookes, and Nancy Kanwisher 467

33 Roles of Visual Area MT in Depth Perception Gregory C. DeAngelis 483

34 Multisensory Integration for Heading Perception in Macaque Visual Cortex Dora E. Angelaki, Yong Gu, and Gregory C. DeAngelis 499

35 Visual Stability during Saccadic Eye Movements Concetta Morrone and David Burr 511

36 Optimal Estimation in Sensory Systems Eero P. Simoncelli 525

V MOTOR SYSTEMS

Introduction Scott T. Grafton and Emilio Bizzi 539

37 Neurobiology of Coordinate Transformations Emilio Bizzi and Ferdinando A. Mussa-Ivaldi 541

38 Basal Ganglia and Cerebellar Circuits with the Cerebral Cortex Richard P. Dum and Peter L. Strick 553

39 The Basal Ganglia and Cognition Ann M. Graybiel and Jonathan W. Mink 565

40 Computational Neuroanatomy of Voluntary Motor Control Reza Shadmehr and John W. Krakauer 587

41 Forward Models and State Estimation in Posterior Parietal Cortex Grant H. Mulliken and Richard A. Andersen 599

42 Parallels between Sensory and Motor Information Processing Emanuel Todorov 613

43 The Mirror Neuron System: A Motor-Based Mechanism for Action and Intention Understanding Giacomo Rizzolatti, Leonardo Fogassi, and Vittorio Gallese 625

44 Relative Hierarchies and the Representation of Action Scott T. Grafton, L. Aziz-Zadeh, and R. B. Ivry 641

VI MEMORY

Introduction Daniel L. Schacter 655

45 Comparative Analysis of the Cortical Afferents, Intrinsic Projections, and Interconnections of the Parahippocampal Region in Monkeys and Rats Wendy A. Suzuki 659

46 Medial Temporal Lobe Function and Human Memory Yael Shrager and Larry R. Squire 675

47 Reconsolidation: A Possible Bridge between Cognitive and Neuroscientific Views of Memory Karim Nader 691

48 The Dynamic Interplay between Cognitive Control and Memory Elizabeth A. Race, Brice A. Kuhl, David Badre, and Anthony D. Wagner 705

49 Phases of Influence: How Emotion Modulates the Formation and Retrieval of Declarative Memories Elizabeth A. Kensinger 725

50 Individual Differences in the Engagement of the Cortex during an Episodic Memory Task Michael B. Miller 739

51 Constructive Memory and the Simulation of Future Events Daniel L. Schacter, Donna Rose Addis, and Randy L. Buckner 751

VII LANGUAGE

Introduction Alfonso Caramazza 765

52 The Cortical Organization of Phonological Processing Gregory Hickok 767

53 Morphological Processes in Language Production Kevin A. Shapiro and Alfonso Caramazza 777

54 Ventral and Dorsal Contributions to Word Reading Laurent Cohen and Stanislas Dehaene 789

55 The Neural Basis of Syntactic Processing David Caplan 805

56 Semantic Unification Peter Hagoort, Giosuè Baggio, and Roel M. Willems 819

57 Early Language Acquisition: Neural Substrates and Theoretical Models Patricia K. Kuhl 837

58 Genetics of Language Franck Ramus and Simon E. Fisher 855

59 The Biology and Evolution of Language: "Deep Homology" and the Evolution of Innovation W. Tecumseh Fitch 873

VIII THE EMOTIONAL AND SOCIAL BRAIN

Introduction Todd F. Heatherton and Joseph E. LeDoux 887

60 Ontogeny of Infant Fear Learning and the Amygdala Regina M. Sullivan, Stephanie Moriceau, Charlis Raineki, and Tania L. Roth 889

61 Emotional Reaction and Action: From Threat Processing to Goal-Directed Behavior Joseph E. LeDoux, Daniela Schiller, and Christopher Cain 905

62 Interactions of Emotion and Attention in Perception Patrik Vuilleumier and Tobias Brosch 925

63 Context Effects and the Amygdala Paul J. Whalen and F. Caroline Davis 935

64 Neurogenetic Studies of Variability in Human Emotion
Ahmad R. Hariri 945

65 Components of a Social Brain Jason P. Mitchell and
Todd F. Heatherton 953

66 The Neural Basis of Emotion Regulation: Making Emotion Work for You and
Not Against You Jennifer S. Beer 961

67 Sharing the Emotions of Others: The Neural Bases of Empathy
Tania Singer and Susanne Leiberg 973

68 The Cognitive Neuroscience of Moral Judgment Joshua D. Greene 987

IX HIGHER COGNITIVE FUNCTIONS

Introduction Elizabeth A. Phelps 1003

69 Prefrontal Substrate of Human Relational Reasoning
Barbara J. Knowlton and Keith J. Holyoak 1005

70 Decision Making and Prefrontal Executive Function
Christopher Summerfield and Etienne Koechlin 1019

71 Circuits in Mind: The Neural Foundations for Object Concepts
Alex Martin 1031

72 Semantic Cognition: Its Nature, Its Development, and Its Neural Basis
James L. McClelland, Timothy T. Rogers, Karalyn Patterson,
Katia Dilkina, and Matthew Lambon Ralph 1047

73 Two Views of Brain Function Marcus E. Raichle 1067

74 The Neuroeconomics of Simple Goal-Directed Choice (Circa 2008)
Antonio Rangel 1075

75 Neuroeconomics and the Study of Valuation Paul W. Glimcher 1085

76 Emotion and Decision Making Elizabeth A. Phelps and
Mauricio R. Delgado 1093

X CONSCIOUSNESS

Introduction Christof Koch 1107

77 Comparing the Major Theories of Consciousness Ned Block 1111

78 Recovery of Consciousness after Brain Injury: An Integrative Research Paradigm
for the Cognitive Neuroscience of Consciousness
Nicholas D. Schiff 1123

79 The Neurobiology of Consciousness Christof Koch 1137

80 Visual Awareness Geraint Rees 1151

81 The Role of Feedback in Visual Attention and Awareness
Stephen L. Macknik and Susana Martinez-Conde 1165

82 Emotion and Consciousness Michael Koenigs and
Ralph Adolphs 1181

83 Volition and the Function of Consciousness Hakwan Lau 1191

84 Toward a Theory of Consciousness Giulio Tononi and
David Balduzzi 1201

XI PERSPECTIVES

85 Mapping Cognitive Neuroscience: Two-Dimensional Perspectives on Twenty Years of Cognitive Neuroscience Research John T. Bruer 1221

86 Reflections on the Cognitive Neuroscience of Language
Sheila E. Blumstein 1235

87 Why the Imagery Debate Won't Go Away Stephen M. Kosslyn,
William L. Thompson, and Giorgio Ganis 1241

88 Looking Toward the Future: Perspectives on Examining the Architecture and Function of the Human Brain as a Complex System Michael S. Gazzaniga,
Karl W. Doron, and Chadd M. Funk 1247

89 The Landscape of Cognitive Neuroscience: Challenges, Rewards, and New Perspectives Elissa M. Aminoff, Daniela Balslev, Paola Borroni,
Ronald E. Bryan, Elizabeth F. Chua, Jasmin Cloutier, Emily S. Cross,
Trafton Drew, Chadd M. Funk, Ricardo Gil-da-Costa, Scott A. Guerin,
Julie L. Hall, Kerry E. Jordan, Ayelet N. Landau, Istvan Molnar-Szakacs,
Leila Montaser-Kouhsari, Jonas K. Olofsson, Susanne Quadflieg,
Leah H. Somerville, Jocelyn L. Sy, Lucina Q. Uddin, and
Makiko Yamada 1255

Contributors 1263

Index 1269

PREFACE

It has been 20 years since we first met in Squaw Valley to assess the state of cognitive neuroscience. We have held this meeting three times before and each meeting had its own signature. When the first meeting concluded, we knew we had a vibrant, young field on our hands. As the years passed, our knowledge deepened and slowly and gradually new ideas emerged. With the fourth meeting cognitive neuroscience is busting out all over. Fundamental stances are changing and new ideas are emerging. Everything from the view that individual neurons change their functional role through time to claims that our moral decisions can be tracked in the brain are indicants of the range and excitement of cognitive neuroscience. Fresh air sweeps in and reinvigorates us on the view that we will some day figure out how the brain works its magic and produces the human mind.

It is always in the first two sessions that one finds the contrast in approaches to studying the brain so markedly different. The development and evolution section talks about a dynamic growth pattern that becomes specific and fixed. At the same time those interested in plasticity see the neuronal systems always changing and the dynamics seen in development as continuing for the life of the brain. In our most recent meeting the reports on brain plasticity were more bold than ever before.

The attention session featured a new emphasis on the interactions between reafferent top-down and feed-forward, bottom-up attentional processes. Benefiting from ever-impressive technological advances, the elucidation of attentional mechanisms is proceeding at a dizzying pace. It is refreshing to note that in addition to providing a more comprehensive picture of attentional processes, this exciting new empirical evidence has also verified many central tenets of some of the most longstanding and influential theories in the cognitive neuroscience of attention.

In the motor session, the boundaries of the motor system continued to be pushed further into the realm of cognition. Some have demonstrated the existence of motor-related areas in the parietal lobe that are involved in representing goals of oneself and of others, providing a link for how we may intuitively translate the actions of others into a model of their mental processes. Separate research indicates that areas once thought to have only motor roles actually contain circuits related to executive and limbic function, further complicating the distinction between cognition and motor processes. Perhaps above all else, the work of the motor section suggests that we might be wise to relieve ourselves of the need to make stark distinctions between these two phenomena, at least in higher primates.

Memory research is, paradoxically, providing great insight into how humans imagine future events. Moreover, new models of reconsolidation and retrieval are emerging, and exciting evidence of individual differences in cortical activation patterns during episodical retrieval are forcing a careful reevaluation of central tenets of functional imaging analysis.

The perception session demonstrated the vast potential of Bayesian modeling to provide beneficial descriptions of how the brain performs various functions. But Bayesian modeling

does not hold a monopoly; other theoretical and methodological pioneers are dramatically enhancing our understanding of vision, quite literally from the level of the retina to large-scale networks that connect distributed regions of the cortex. Keeping pace with advances in vision science are exciting findings across the modalities of audition, olfaction, and vestibular function.

Next, we turned our attention to language (and secretly hoped that a few lectures would pass without mention of the word "Bayesian"). The understanding of specific components of language processing is expanding, while exciting parallel studies are examining the genes that may wire our brain in a way that enables language acquisition. But even as we move toward an understanding of how genes and experience sculpt the human brain into a speaking device, the question of who exactly is doing the speaking arises. Fittingly, we transitioned into the session on executive function, where we came across a surprising answer to this question. There does not appear to be a need for a "top" in top-down control; instead, various regions for self-regulation and cognitive control have been identified and their interactions have been modeled in ways that leave the mythical homunculus homeless. As if this were not profound enough, we also learned about an exciting new characterization of resting brain activity, a remarkable advance that has too many implications to list.

Over a week removed from our introduction to theory of mind in the motor session, the emotion and social neuroscience section further demystified the rapidly expanding science of the social brain. New ideas on how our emotions and sense of self inform the ways in which we think about and reflexively understand others continue to evolve, while evidence for the genetic basis of individual variation in affect and, astoundingly, for differences in BOLD activity related to this genetic variation has further shaped the current models of how the emotional brain develops and operates.

As it always does, the conference ended with a bang, featuring two days' worth of lively discussion on the topic of consciousness. An exciting novel mechanism for how the brain generates the baseline activity necessary to sustain conscious experience was complemented by a bold theoretical attempt to make the problem of qualia a bit more tractable. Between these extremes, others reported suggestive new evidence about the neural basis of visual conscious experience. This session also served as a reminder of how far we had come during those three weeks in Squaw Valley, as topics such as action, emotion, language, and executive function reemerged in the context of examining how such varied processing contributes to the content of conscious experience.

After three weeks of such intense stimulation, it is a testament to the amazing progress unveiled that one somehow leaves Tahoe reinvigorated and enthusiastic to get back into the lab. The past 20 years have seen advances that we could never have anticipated, and, incredibly, the next five or ten hold the potential to continue this exponential progress. The Summer Institute at Tahoe reveals simultaneously the exciting new ideas in the field and the bright minds that are vigorously attacking the persisting mysteries of cognitive neuroscience. It also exposes a talented group of graduate and postdoctoral students to the wonderful breadth and depth of the field. Scanning the room of eager young minds hanging onto the words of the various leaders of the field is truly a sight to behold. It is exhilarating to witness the handing of the baton from one generation to the next. One can only pause, take it all in, smile, and then get back to paying attention to the lecture because the next big idea presented might knock you right out of your seat!

Needless to say, complex events and publications like this work well only if there are dedicated people involved. First, the MIT Press continues to be exceptionally supportive in carrying out high-quality production in a timely manner. Once again my daughter, Marin Gazzaniga, managed the ebb and flow of the manuscripts, playing both good cop and bad cop as the manuscripts moved between authors, section editors, and ultimately the publisher. Marin is a brilliant playwright, actress, and writer in her own right, and all of those skills are required in herding academics to a common goal.

The actual event at Lake Tahoe was managed from the beginning by my assistant, Jayne Rosenblatt. She is always good humored and incredibly dedicated and runs complex events seemingly effortlessly. Finally, these books don't just happen. Peggy Gordon brings it all together into print with a steady hand and professionalism.

Warm thanks and congratulations to all. We will see you all again in five years.

Michael S. Gazzaniga
The Sage Center for the Study of Mind
University of California, Santa Barbara

I

DEVELOPMENT AND EVOLUTION

Chapter 1 RAKIC, ARELLANO, AND BREUNIG 7

2 KOSTOVIĆ AND JUDAŠ 29

3 PREUSS 49

4 CHALUPA AND HUBERMAN 67

5 BUNGE, MACKEY, AND WHITAKER 73

Introduction

PASKO RAKIC AND LEO M. CHALUPA

In the 15 years since the first edition of *The Cognitive Neurosciences* we have witnessed immense advances in the understanding of the intricacies of human cognitive abilities. As evident from the many articles in this as well as previous editions of this popular reference book, the progress that has been made can to a large extent be linked with studies of the cerebral cortex using very sophisticated noninvasive imaging methods in human subjects. These methods allow examination of human-specific cognitive functions directly in living people as they develop, as they are perturbed, and as they decline (e.g., Gazzaniga, 2008). It is thus somewhat paradoxical that during this very same time period the major advances in our understanding of the cellular and molecular mechanisms of cortical development and the models of evolutionary elaborations have derived almost entirely from studies of rodent brains.

Although basic principles of cortical development are probably similar in all species, the modifications of developmental events during evolution produce not only quantitative changes (e.g., the number of neurons; expansion in surface, timing, and sequence of cellular events; increase in number of synapses, etc.), but also many qualitative changes (e.g., the elaboration of new types of neurons and glia, and most importantly additions of novel specialized cytoarchitectonic areas associated with corresponding new pathways and patterns of connectivity). It has become evident that even essential genes that are responsible for survival of the species give different phenotypes in the mouse and human (Liao & Zhang, 2008). Furthermore, the timing and duration of cell genesis, the composition of the ventricular zone, and the ratios of cell proliferation, versus programmed cell death occurring in the enlarged subventricular and marginal zones of the embryonic cerebral cortex, suggest not only a slower development, but also

expanded, diversified, and novel roles of these transitional layers in primates including human (Bystron, Blakemore, & Rakic, 2008). It is for this reason that the first three chapters in this volume are dedicated to the development and evolution of the cerebral cortex with a particular emphasis on humans and nonhuman primates.

The first chapter, by Rakic, Arellano, and Breunig, is dedicated to the prenatal development of the primate neocortex. It emphasizes the developmental features that are prominent and essential for the formation of the large and convoluted cerebral cortex. For example, there are marked species-specific differences in the timing and sequence of divergence of neural stem and radial glial cell lines as well as in the levels of their differentiation and longevity. There are also subclasses of neural stem cells that produce interneurons for the neocortex as well as for the association thalamic nuclei not detectable in rodents (Letinic & Rakic, 2001; Letinic, Zonku, & Rakic, 2002). These neurons may be involved in human-specific language and cognitive functions that do not exist in nonprimate species.

The second chapter, by Kostović and Judaš, describes the early development of neuronal circuitry of the human prefrontal cortex, which is most elaborated and enlarged in humans, and arguably does not exist in nonprimate species. These studies are helped by the increase in resolution of the MRI to the degree that one can visualize normal and possible abnormal columns in the human fetal neocortex (e.g., McKinstry et al., 2002).

The third chapter, by Preuss, which deals with evolutionary aspects of cortical development in the hominoids, provides a compelling account of recent evidence documenting unique features in the organization of the human brain. These new insights have been derived by using new technologies, in combination with more established techniques, to assess different aspects of the relation between structure and function in the brain, ranging from neuronal morphology to connectional pathways.

The fourth chapter, by Chalupa and Huberman, is concerned with unraveling the role of neuronal activity in the formation of eye-specific connections in nonhuman primates. It deals with the development, competition, and plasticity of the projections of the two eyes to the lateral geniculate nucleus and the formation of the ocular dominance columns in the primary visual cortex. Their work challenges the widely held notion that neuronal activity, in particular the retinal waves of activity, plays an instructional role in the formation of eye-specific retinogeniculate projections.

The fifth chapter, by Bunge, Mackey, and Whitaker, deals with human brain changes underlying improved cognitive abilities during childhood and adolescence, with a particular emphasis on fluid reasoning. Focusing primarily on prefrontal and parietal cortices and using the most advanced neuroimaging methods and conceptual approaches, this chapter aptly demonstrates the power of the developmental approach in linking brain mechanisms with higher cognitive functions.

Collectively, these studies may help in understanding the biological bases of the high level of cognitive ability that is achieved during primate evolution culminating in humans. However, from a practical perspective, the findings obtained from studies on human and nonhuman primates may be essential for the design of psychiatric drug therapies, since, for example, the capacity for regeneration has diminished during vertebrate evolution, and the absence of neurogenesis in the primate cerebral cortex (Bhardwaj et al., 2006) indicates that overcoming the brain's resistance to the acquisition of functionally competent new neurons will require an understanding of why neurogenesis ceases at the end of specific developmental time windows and why there are regional variations in this phenomenon (Rakic, 2002, 2006). Another difference is the existence of distinct types of interneurons in the human brain that are not detectable in rodent species (e.g., DeFelipe, Alonso-Nanclares, & Arellano, 2002). In addition, a subclass of interneurons of the thalamic association nuclei that originate in the ganglionic eminence are not detectable in rodents (Letinic & Rakic, 2001). Likewise, unlike in rodents, in which interneurons arise from the ganglionic eminences, in primates these originate in large numbers in the enlarged subventricular zone (Letinic et al., 2002; Petanjek, Dujmovic, Kostovic, & Esclapez, 2008). These neurons may be involved in human-specific disorders such as schizophrenia that do not occur spontaneously in nonprimate species. Thus modifications in the expression pattern of transcription factors in the human forebrain may underlie species-specific programs for the generation of specific classes of cortical neurons that may be differentially affected in genetic and acquired neurological disorders (Lewis, 2000). These novel evolutionary traits may be more vulnerable to genetic mutations and environmental insults, and could be implicated in disorders of higher brain functions, such as autism, developmental dyslexia, Alzheimer's disease, and schizophrenia. Designs of new drugs and replacement therapies need to take into consideration these species-specific distinctions.

REFERENCES

Bhardwaj, R. D., Curtis, M. A., Spalding, K. L., Buchholz, B. A., Fink, D., Bjork-Eriksson, T., et al. (2006). Neocortical neurogenesis in humans is restricted to development. *Proc. Natl. Acad. Sci. USA*, *103*, 12564–12568.

Bystron, I., Blakemore, C., & Rakic, P. (2008). Development of the human cerebral cortex: Boulder Committee revisited. *Nat. Rev. Neurosci.*, *9*, 110–122.

DeFelipe, J., Alonso-Nanclares, L., & Arellano, J. I. (2002). Microstructure of the neocortex: Comparative aspects. *J. Neurocytol.*, *31*, 299–316.

Gazzaniga, M. S. (2008). *Human: The science behind what makes us unique*. New York: HarperCollins.

Letinic, K., & Rakic, P. (2001). Telencephalic origin of human thalamic GABAergic neurons. *Nat. Neurosci.*, *4*, 931–936.

Letinic, K., Zoncu, R., & Rakic, P. (2002). Origin of GABAergic neurons in the human neocortex. *Nature*, *417*, 645–649.

Lewis, D. A. (2000). GABAergic local circuit neurons and prefrontal cortical dysfunction in schizophrenia. *Brain Res. Brain Res. Rev.*, *31*, 270–276.

Liao, B. Y., & Zhang, J. (2008). Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc. Natl. Acad. Sci. USA*, *105*, 6987–6992.

McKinstry, R. C., Mathur, A., Miller, J. H., Ozcan, A., Snyder, A. Z., Schefft, G. L., et al. (2002). Radial organization of developing preterm human cerebral cortex revealed by noninvasive water diffusion anisotropy *MRI. Cereb. Cortex*, *12*, 1237–1243.

Petanjek, Z., Dujmovic, A., Kostovic, I., & Esclapez, M. (2008). Distinct origin of GABAergic neurons in forebrain of man, nonhuman primates and lower mammals. *Collegium Anthropologicum*, *32*, Suppl. 1, 9–17.

Rakic, P. (2002). Neurogenesis in adult primate neocortex: An evaluation of the evidence. *Nat. Rev. Neurosci.*, *3*, 65–71.

Rakic, P. (2006). Neuroscience: No more cortical neurons for you. *Science*, *313*, 928–929.

1 Development of the Primate Cerebral Cortex

PASKO RAKIC, JON I. ARELLANO, AND JOSHUA BREUNIG

ABSTRACT The cerebral cortex is the crowning achievement of evolution and the biological substrate of human cognitive abilities. Although the basic principles of cortical development in all mammals are similar, the modifications of developmental events during evolution produce not only quantitative but also qualitative changes. Human cerebral cortex, as in the other species, is organized as a map in which specific cell classes are positioned into a radial, laminar, and areal array that depends on the sequential production and phenotypic specification of those cells and the directed migration from their place of origin to their distant final destination. The long and curvilinear migration pathways in the fetal human cerebrum depend critically on the stable radial glial scaffolding. After neurons assume their proper areal and laminar position, they attach locally and develop numerous proximal and long-distance connections that involve specific adhesion molecules, neurotransmitters, and receptors. However, the final pattern of synaptic connections is selected through functional validation and selective elimination of the initially overproduced neurons, axons, and synapses. In this review, the development of the cerebral cortex is described in the context of the radial unit hypothesis, the postulate of an embryonic protomap, and the concept of competitive neural interactions that ultimately create a substrate for the highest cognitive functions.

There is probably no disagreement among biologists that the cerebral cortex is the part of the brain that most distinctively sets us apart from any other species and that the principles governing its development may hold the key to explaining our cognitive capacity, intelligence, and creativity (e.g., Gazzaniga, 2008). Perhaps the most prominent feature of the cerebral cortex in all species, and particularly in primates, is its parcellation into distinct laminar, radial, and areal domains (Eccles, 1984; Mountcastle, 1997; Goldman-Rakic, 1987; Rakic, 1988; Szentagothai, 1978). Although the surface of the neocortex has expanded a thousandfold during phylogeny, its thickness and its basic cytoarchitectonic organization appear to be changed comparatively less. However, this morphological similarity of cortical architecture in histological sections in all mammals may be misleading, since it might invite us to think in terms of a canonical cortex that only varies in size in accordance with the variations of body mass of the different mammalian species. Conversely, our current knowledge indicates that there are significant qualitative and quantitative changes in the structure of the neocortex between species, in the size of cells, in the proportion of neurons to glia, in the ratio of excitatory projection cells to inhibitory interneurons, in the appearance of new types of neurons, in the specific organization of connections, and, most importantly, in the addition of novel highly specialized cortical areas associated with correspondingly new axonal pathways and patterns of synaptic connectivity that can certainly have a profound effect on the functional capacity of the cerebral cortex. In spite of these differences, the small size, high fertility rate, and low cost of maintenance have converted mice into an unexcelled model for experimental research on neuroscience and particularly on basic cortical organization (Rakic, 2000).

The study of the cortical development of mammals suggests that even small differences in the timing and duration of the genesis of neural cells and changes in the composition and the ratios of cell proliferation and programmed cell death in the transient embryonic zones of the developing forebrain can be responsible for the evolutionary expansion of the neocortex and the concomitant appearance of novel functions in cognitive processing of information (Bystron, Blakemore, & Rakic, 2008). Although genetically humans are surprisingly similar to other mammals, the uniqueness of the human cognitive potential, which is an output of cortical function, must have some genetic basis. Since the molecular structure and function of neurotransmitters, receptors, and ion channels do not change substantially over the phylogenetic scale, the secret to the success of *Homo sapiens* is probably mainly due to an increased number of neurons, more elaborated connections, functional specialization, and introduction of new cortical areas. Even the so-called essential genes, which are considered responsible for survival of the individual, give different phenotypes in different species, and about 20% of the mouse orthologs of human-essential genes are nonessential in mice (Liao & Zhang, 2008). It is therefore apparent that neither the genetic, the cellular, nor most importantly, the circuitry basis of human cortical uniqueness can be deciphered by studying exclusively rodents, much as one cannot expect to understand the origin

PASKO RAKIC, JON I. ARELLANO, and JOSHUA BREUNIG Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut

of the mushroom body of insects by studying the primate association cortex. However, for logistic, ethical, and financial reasons, the easiest approach to understanding general mammalian features is to study the mouse, and the closest we can come to elucidating primate-specific cortical development with the modern molecular and cell biological methods is to analyze its development in the monkey, and specifically in Old World monkeys such as the macaque, which are much more similar to the human species than their relatives the New World monkeys. For the preceding reasons the present review is based on the comparative studies of neocortical development in the mouse, macaque, and human. It is limited to the early developmental events that lead to the formation of cellular constituents and their basic connectivity. The final tuning of synaptic connections and their modifications by experience is described in other communications in this volume.

Onset and span of cortical neurogenesis

The mammalian cerebral cortex is a laminated structure composed of a bewildering diversity of neurons arranged in distinct cytoarchitectonic fields. It is well established, but nevertheless always fascinating, that none of these neurons are generated in the cortex itself. Classical studies of neurogenesis based on the distribution of mitotic figures and deployment of migrating neurons in the cerebral wall in the human fetus suggested that all cortical neurons in the human are likely to be generated mostly before birth (e.g., Poliakov, 1959, 1965; reviewed in Sidman & Rakic, 1973, 1982; Rakic 2002; Bystron et al., 2008). However, the precise data on the onset and termination of corticogenesis could not be established with the classical histological techniques alone (e.g., Conel, 1939), a limitation that was overlooked with the introduction of tritiated thymidine and BrdU incorporation as markers of DNA synthesis to label dividing cells (Breunig, Arellano, Macklis, & Rakic, 2007). These new methods allowed a systematic study of neuronogenesis, and it was established that the neocortex receives the first neurons at ~E10 in mice, and ~E33 in macaque and human (reviewed in Bystron et al., 2008). These studies also showed that there does not exist a pattern between species for the timing of neuronal genesis, and while in mice neurons are generated in the second half of gestation (between E10 and the end of gestation at E18.5), in the macaque monkey and human, neurons are produced much earlier, mostly during the middle of gestation in monkeys and during the first half in humans (between E33 and E100 in monkeys and E33 and E120 in humans, figure 1.5) (Rakic, 1974, 1988, 2002; Rakic & Sidman, 1968; Sidman & Rakic, 1973, 1982). This early neocortical genesis stands in contrast to that of the cerebellum, olfactory bulb, and hippocampus, which in the mouse terminate their development postnatally, and in primates including human continue to add neurons well after birth (Rakic, 1973; Rakic & Nowakowski, 1981; Kornack & Rakic 1999, 2001b).

In spite of comprehensive search in the adult monkey neocortex, no additional neurons are added during the animal's 30-year life span (Rakic, 1985; Kornack & Rakic 2001a; Koketsu, Mikami, Miyamoto, & Hisatsune, 2003). The claim of new neurons arriving to the prefrontal, parietal, and temporal association cortices in adult macaques could not be confirmed (reviewed in Breunig, Arellano, et al., 2007; Rakic, 2002, 2006), and additional techniques such as the analysis of incorporation of C14 has provided further evidence that all neurons of the human neocortex are generated before birth (Bhardwaj et al., 2006), indicating that for the highest cognitive functions we use and depend on the same set of neurons throughout our entire life span (Rakic, 1985, 2006).

Place of origin

The presence of mitotic figures near the lumen of the cerebral cavity of the embryonic human cerebrum and their paucity and/or absence in the cortical plate itself led to the hypothesis that cortical neurons are produced in the germinal matrix situated at the ventricular surface (His, 1904), an idea that was substantiated by the labeling of dividing cells in mice (Angevine & Sidman, 1961), monkeys (Rakic, 1974), and humans (Rakic & Sidman, 1968; Letinic, Zoncu, & Rakic, 2002). In the last decade, studies in rodents have established two sources of origin of cortical neurons: the germinal regions of the dorsal telencephalon give rise to cortical projection neurons that migrate radially to their final position in the cortex, while the ganglionic eminences in the ventral telencephalon give rise to virtually all GABAergic cortical interneurons (e.g., Lavdas, Grigoriou, Pachnis, & Parnevelas, 1999; Marin & Rubenstein, 2001), which migrate tangentially via the marginal and intermediate zones to the cortex (Ang, Haydar, Gluncic, & Rakic, 2003). This model has been observed in rodents and carnivores (ferrets; Anderson, Kaznowski, Horn, Rubenstein, & McConnell, 2002; Poluch & Juliano, 2007), but also in avians such as chickens (Cobos, Puelles, & Martinez, 2001), and a similar population of tangentially migrating neurons has been observed in the human using classical histological material (Rakic, 1975). However, in what appears to be a significant species-specific difference, neocortical inhibitory interneurons in primates, including human, originate not only in the ganglionic eminences in the ventral telencephalon, but also in the ventricular zone and in the enormously enlarged subventricular zone of the dorsal telencephalon (Letinic et al., 2002; Rakic & Zecevic, 2003; Petanjek, Berger, & Esclapez, 2008; Petanjek, Dujmovic, Kostovic, & Esclapez, 2009) and quantitative data from Letinic and

colleagues (2002) indicated that about two-thirds of the cortical GABAergic interneurons might have a dorsal telencephalic origin in humans.

Proliferative cells in the ventricular zone are organized as a pseudostratified epithelium in which precursor cells divide asynchronously; their nuclei move away from the ventricular surface to replicate their DNA and then move back to the surface to undergo another mitotic cycle (reviewed in Sidman & Rakic, 1973; Rakic, 1988). Early silver impregnation methods revealed a distinct population of elongated, nonneuronal cells in the fetal human brain, which were initially called epithelial cells or fetal glia (Rakic, 2003). Later electron microscopic and immunohistochemical studies using glial acidic fibrillary protein (GFAP) in human and nonhuman primates confirmed their glial nature (Levitt, Cooper, & Rakic, 1981) and justified the use of the term "radial glial cells" (RGC; Rakic, 1988). The distal end-feet of the radial glial cells form the pial surface of the fetal cerebrum (Rakic, 1972). A variety of antigens, such as vimentine, brain-lipid-binding protein (BLBP), the astrocyte specific glutamate transporters (GLAST), and RC1, RC2 have been used to characterize their glial nature (e.g., Bystron et al., 2008).

Radial glia is particularly prominent in the embryonic primate cerebrum, where a subset of GFAP positive cells transiently stops dividing (Schmechel & Rakic, 1979). However, it has become now clear that the RGC can generate both neuronal progenitors and neurons (Cameron & Rakic, 1991; Malatesta, Hartfuss, & Gotz, 2000; Malatesta et al., 2003; Hartfuss, Galli, Heins, & Gotz, 2001; Noctor, Flint, Weissman, Dammerman, & Kriegstein, 2001; Noctor et al., 2002; Tamamaki, Nakamura, Okamoto, & Kaneko, 2001; Fishell & Kriegstein, 2003; Tramontin, Garcia-Verdugo, Lim, & Alvarez-Buylla, 2003, Gal et al., 2006). The dividing RGC generate successive neuronal clones that migrate along the elongated parental process into the cortical plate or populate the subventricular zone (SVZ) where they divide again before entering a postmitotic state and migrate to the overlying cortex. In addition, they generate intermediate progenitors that continue to divide (Levitt et al., 1981; Gal et al., 2006). The SVZ was previously thought to generate mainly glia (Altman & Bayer, 1990). However, it is now well established that this zone is multipotential and also generates projection neurons and interneurons, as well as various types of glial cells in all species studied, although the proportion is different between species (Letinic et al., 2002; reviewed in Bystron et al., 2008). Thus, in rodents and in primates, early divergence of basic cell types has been revealed using the retroviral gene transfer method, which enables the study of cell lineages in the developing mammalian telencephalon (Luskin, Pearlman, & Sanes, 1988; Cameron & Rakic, 1991; Kornack & Rakic, 1995). It appears that in the primate, the radial glial mother cell gives rise to a daughter cell which either directly or as a dedicated neuronal progenitor goes through several rounds of division to produce bipolar migrating neurons that will migrate up the radial process of the mother cell (Rakic, 2003). Current state-of the-art methods including in utero electroporation and conditional mouse genetics, which both allow for rapid and powerful gain- and loss-of-function studies, are beginning to unravel the complex molecular interplay between these cell types in the developing cortex. A complete compendium of newly identified signaling pathways is beyond the scope of this chapter, but a few findings deserve mention. For example, it has been demonstrated that the protein Numb is a crucial player in maintaining the adhesiveness of radial glia in the ventricular zone (VZ), preventing premature detachment and subsequent astrogliogenesis (Rasin et al., 2007). In contrast, Notch functions cell-autonomously to maintain the radial glial cell fate while the proneural genes antagonize Notch signaling to promote neuronal differentiation and subsequent migration (Breunig, Silbereis, Vaccarino, Sestan, & Rakic, 2007; Mizutani, Yoon, Dang, Tokunaga, & Gaiano, 2007; Shimojo, Ohtsuka, & Kageyama, 2008; Ge et al., 2006). More precisely, in an example of the exquisite balance of structure and function, it has been shown that the daughter neuronal cell stimulates Notch signaling in the radial glial mother cell to maintain the neurogenic VZ niche and migratory scaffold (Yoon et al., 2008).

Transient embryonic zones

The formation of the adult cortex is the end product of a series of morphogenetic steps that are initiated in the proliferative epithelium at the surface of the lateral cerebral ventricles. Initial cellular events, such as the proliferation, migration, aggregation, and selective death of some of the generated cells, as well as the subsequent outgrowth of axons and dendrites, and the establishment of neuronal connections, proceed in an orderly fashion in each species according to a species-specific timetable that is regulated by differential gene expression. During embryonic and fetal stages, the telencephalic wall consists of several cellular layers that do not exist in the mature brain. These layers, or zones, were recognized by a committee appointed by the American Association of Anatomists (Boulder Committee, 1970) based on the data from the dissertation on human brain development by the senior author of the present chapter (P.R.). The committee recommendations for the names of the transient embryonic cellular compartments, termed zones, have been adopted as a generic description for fundamental developmental events for the entire vertebrate central nervous system. However, in the past four decades, the development of new techniques, particularly genetic tools for fate mapping and gain and loss of function of targeted genes, have contributed largely to elucidate the

genetic regulation of the developmental processes in various species. In particular, the focus has been set on the patterns of expression of transcription factors that seem to exquisitely influence regional differentiation and regulate broad aspects of mitotic activity, fate choice, migration, and differentiation (Gleeson & Walsh, 2000; Ge et al., 2006; Guillemot, 2007; Mizutani et al., 2007; Molyneaux, Arlotta, Menezes, & Macklis, 2007; Rasin et al., 2007). Recent studies have revealed new types of transient neurons and proliferative cells outside the classical *neuroepithelium*, new routes of cellular migration, and additional cellular compartments (Bielle et al., 2005; Bystron, Rakic, Molnar, & Blakemore, 2006; Carney, Bystron, Lopez-Bendito, & Molnar, 2007; Smart, Dehay, Giroud, Berland, & Kennedy, 2002; Zecevic, Chen, & Filipovic, 2005). As a consequence, a revision of the Boulder Committee model that incorporates that new knowledge has recently been proposed (reviewed in Bystron et al., 2008). In figure 1.1 top, the summary diagram from the original Boulder model is reproduced, and at the bottom the new drawing that incorporates new cell types and cellular zones is shown (Bystron et al., 2008).

Although most of the transient embryonic zones defined by the Boulder Committee were described in the classical literature (e.g., His, 1904), the subplate zone (SP) has been recognized as a separate entity only relatively recently (figure 1.2) (Kostovic & Molliver, 1974; reviewed in Kostovic & Rakic, 1990). This zone consists of early generated neurons scattered among numerous axons, dendrites, glial fibers, and migrating neurons. Although it has been suggested that the subplate zone provides an opportunity for interactions between incoming afferent fibers and early generated neurons, the significance of these transient contacts is not fully understood. Another suggestion was that the subplate zone serves as a cellular substrate for competition among the initial contingent of cortical afferents and that this competition serves to regulate their distribution to appropriate regions of the overlying cortical plate (Rakic, 1976b, 1977; Kostovic & Rakic, 1984; McConnell, Ghosh, & Shatz, 1994).

Subsequent autoradiographic, electron microscopic, and histochemical studies revealed that the axons observed in the subplate zone originate sequentially from the brain stem, basal forebrain, thalamus, and the ipsi- and contralateral cerebral hemispheres (figure 1.2) (Kostovic & Rakic, 1990). More recently, the subplate has been shown to be important for the formation of functional architecture in the cortex such as ocular dominance columns in the visual cortex and also for ensuring the proper formation and strengthening of synapses in this area (Kanold, Kara, Reid, & Shatz, 2003). After a variable and partially overlapping period, these diverse fiber systems enter the cortical plate, the subplate zone disappears, and most of these subplate neurons eventually degenerate, leaving only a vestige of cells scattered throughout the subcortical white matter, which are known as interstitial neurons (Kostovic & Rakic, 1980; Luskin & Shatz, 1985; Chun & Shatz, 1989). A comparison among various species indicates that the size and role of this transient zone increases during mammalian evolution and culminates in parallel with the developing of association areas of the human fetal cortex concomitantly with the enlargement of the cortico-cortical fiber systems (Kostovic & Rakic, 1990; Kostovic & Goldman-Rakic, 1983). The regional differences in the size, pattern, and resolution of the subplate zone correlate also with the pattern and elaboration of cerebral convolutions (Goldman-Rakic & Rakic, 1984). Another transient layer that forms between the neuroepithelium and the pial surface of the dorsal telencephalon that was not mentioned in the Boulder Committee is the preplate that has been widely used and is included in the new schema. Finally, cell divisions outside the classical proliferative zones have been included (Bystron et al., 2008).

Neuronal cell migration

Since all cortical neurons originate near the ventricular surface of the cerebral vesicle, they must all move to their final positions in the cortex, which develops in the outer regions of the cerebral wall, just below the pia. After their last division, postmitotic cells become polarized, extending the leading processes toward the pia and then translocating the nucleus and surrounding cytoplasm within that process (Rakic, 1971, 1972). The extension of the leading process and nuclear translocation are inseparable cellular events. Initially, while the cerebral wall is relatively thin, the tip of the leading process can reach the cortical plate, and the nucleus needs to move only a short distance in both the small rodent cerebrum as well as in the human at comparably early embryonic stages (Sidman & Rakic, 1973; Nadarajah, Alifragis, Wong, & Parnavelas, 2003). However, during the subsequent course of corticogenesis, the cerebral hemispheres enlarge and the length of the migratory pathway increases, particularly in the large primate cerebrum in which, during midgestation, a massive migration of neurons occurs concomitantly with the rapid growth of the cerebral wall. This magnitude of cell movement is perhaps the reason that neuronal cell migration was first observed in human embryos (His, 1874). When this large increase in length occurs, bipolar migrating neurons do not span the entire width of the cerebral wall (reviewed in Sidman & Rakic, 1982; Rakic, 1988, 1990). Subsequent time-lapse imaging studies in the mouse forebrain have shown that this way of migration, termed somal translocation, occurs in all cells irrespective of the length of the migratory pathway (Rakic, 1971, 1972; Nadarajah et al., 2003).

In the early 1970s it was discovered that postmitotic neurons find their way to the cortex by following the

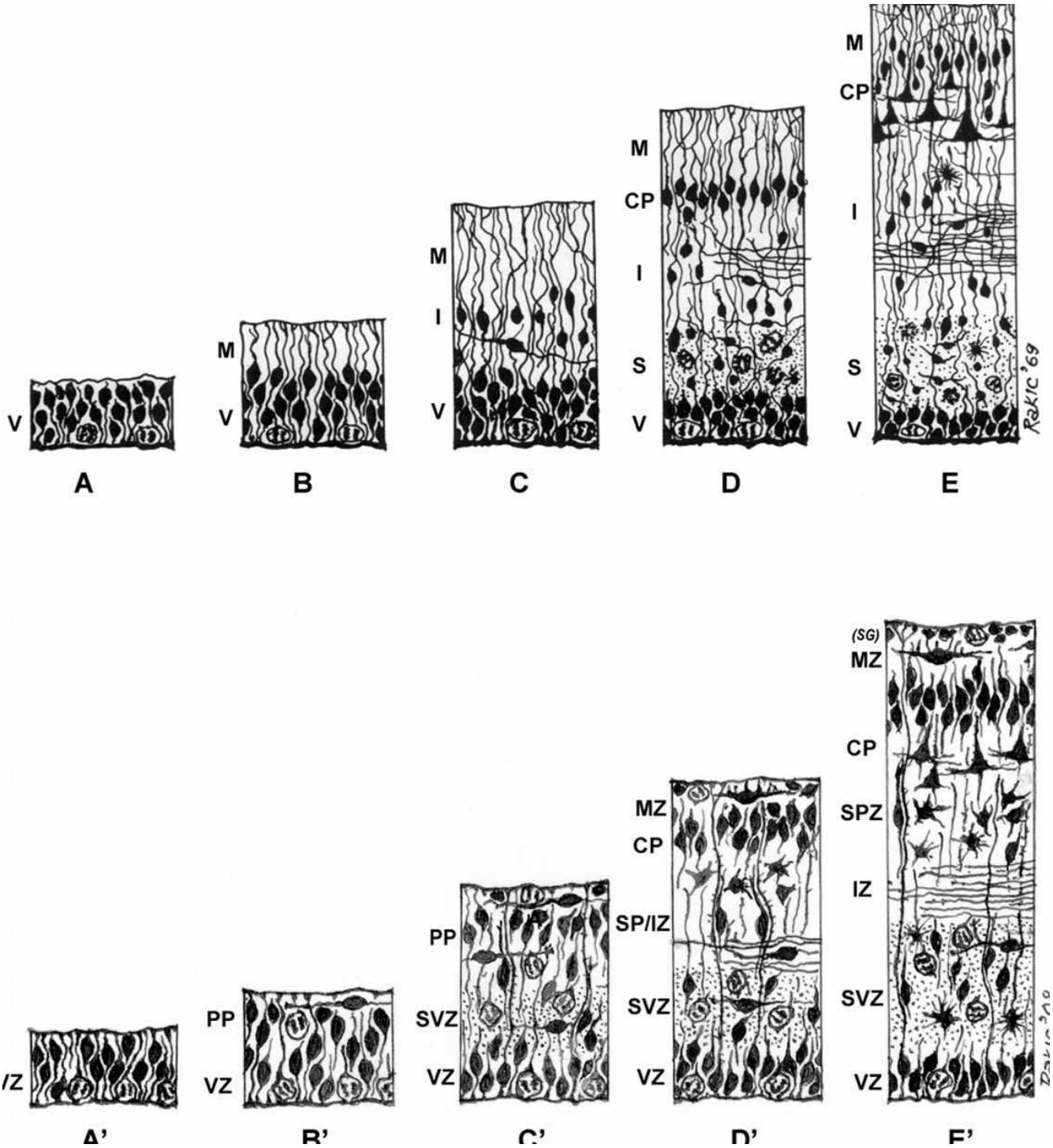


FIGURE 1.1 The original Boulder Committee's (1970) diagram (top) and revised version reprinted from Bystron et al. (2008). Abbreviations in the original diagram (top): V, ventricular zone; M, marginal zone; I, intermediate zone; S, subventricular zone; CP, cortical plate. Abbreviations in the updated diagram (bottom): VZ, ventricular zone; PP, preplate; SP/IZ, subplate/intermediate zone; SVZ, subventricular zone; MZ, marginal zone; CP, cortical plate; SPZ, subplate; IZ, intermediate zone; (SG), subpial granular layer (part of the MZ). The lower panels correspond to the following approximate ages (for the lateral part of the dorsal telencephalon): A′, E30; B′, E31–32; C′, E45; D′, E55; E′, 14 GW.

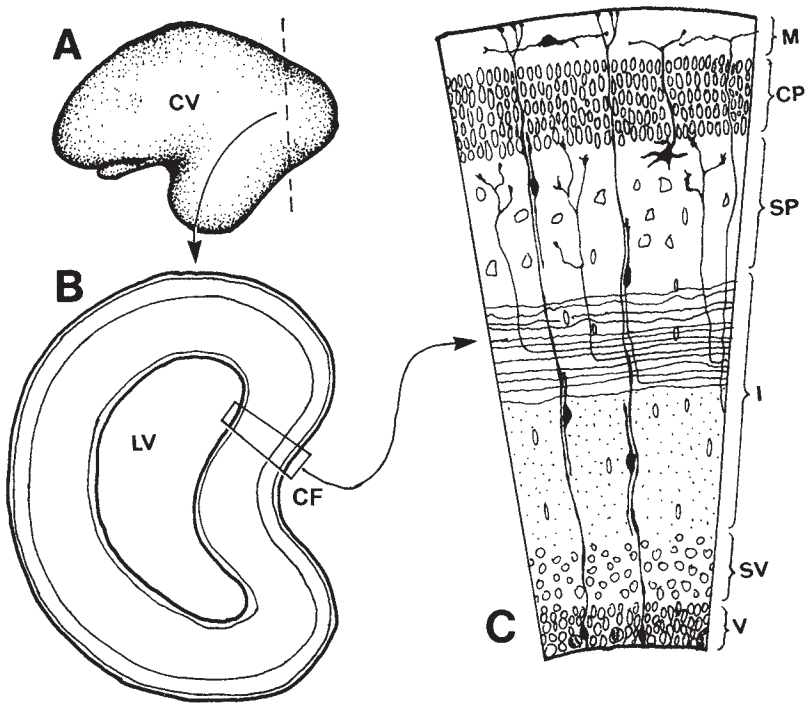


FIGURE 1.2 Cytological organization of the primate cerebral wall during the first half of gestation. (*A*) The cerebral vesicle of 60–65-day-old monkey fetuses is still smooth and lacks the characteristic convolutions that will emerge in the second half of gestation. (*B*) Coronal section across the occipital lobe at the level indicated by a vertical dashed line in *A*. The lateral cerebral ventricle at this age is still relatively large, and only the incipient calcarine fissure (CF) marks the position of the prospective visual cortex. (*C*) A block of the tissue dissected from the upper bank of the calcarine fissure. At this early stage one can recognize six transient embryonic zones from the ventricular surface (bottom) to the pial surface (top): ventricular zone (V); subventricular zone (SV); intermediate zone (I); subplate zone (SP); cortical plate (CP); and marginal zone (M). Note the presence of spindle-shaped migrating neurons moving along the elongated radial glial fibers, which span the full thickness of the cerebral wall. The early afferents originating from the brain stem, thalamus, and other cortical areas invade the cerebral wall and accumulate initially in the subplate zone, where they make transient synapses before entering the overlying cortical plate. (From Rakic, 1995b.)

elongated shafts of radial glial cells (figure 1.3 and Rakic, 1972). Radial glial cells are particularly prominent in primates including human, whose fibers span all the length of the convoluted cerebral hemispheres at the late stages of corticogenesis (Rakic, 1976b; deAzevedo 2003). While moving along the glial surface, migrating neurons remain preferentially attached to curvilinear glial fibers, a finding which suggested a "gliophilic" mode of migration (Rakic, 1985, 1990) that may be mediated by heterotypic adhesion molecules (Rakic, Cameron, & Komuro, 1994). As many as 30 migrating GFAP-negative neurons have been observed migrating along a single GFAP-positive radial glial fascicle in the human forebrain during midgestation (Rakic, 2003).

However, some postmitotic cells do not obey glial constraints and move along tangentially oriented axonal fascicles (e.g., black horizontally oriented cells aligned with thalamic radiation (TR in figure 1.3). We suggested the term "neurophilic" to characterize the mode of migration of this cell class (Rakic, 1985, 1990). Although lateral dispersion of postmitotic neurons was initially observed in Golgi-stained preparations (e.g., figure 1*A* of the report of the Boulder Committee, 1970), it attracted renewed attention after the characterization of the migration through the rostral migratory stream from the postnatal ventricular zone to the olfactory bulb (Menezes & Luskin, 1994; Lois & Alvarez-Buylla, 1994) and from the ganglionic eminence to the dorsal neocortex (De Carlos, Lopez-Mascaraque, & Valverde, 1996; Tamamaki, Fujimori, & Takauji, 1997). Studies in rodents also suggested a more widespread dispersion of clonally related cortical cells (reviewed in Rakic, 1995a; Tan et al., 1998; Reid, Liang, & Walsh, 1995). Although in rodents most of tangentially migrating cells in the dorsal neocortex are inhibitory GABAergic interneurons (reviewed in Marin & Rubenstein, 2001) there are also a number of migrating oligodendrocytes (He, Ingraham, Rising, Goderie, & Temple, 2001).

It should be underscored, however, that clonal analysis in the convoluted primate cortex revealed that the majority of migrating cells, both projection cells and interneurons, obey the radial constraints imposed by the radial glial scaffolding (Kornack & Rakic, 1995; see also Rakic, 2007, and the following section on the radial unit hypothesis). Also,

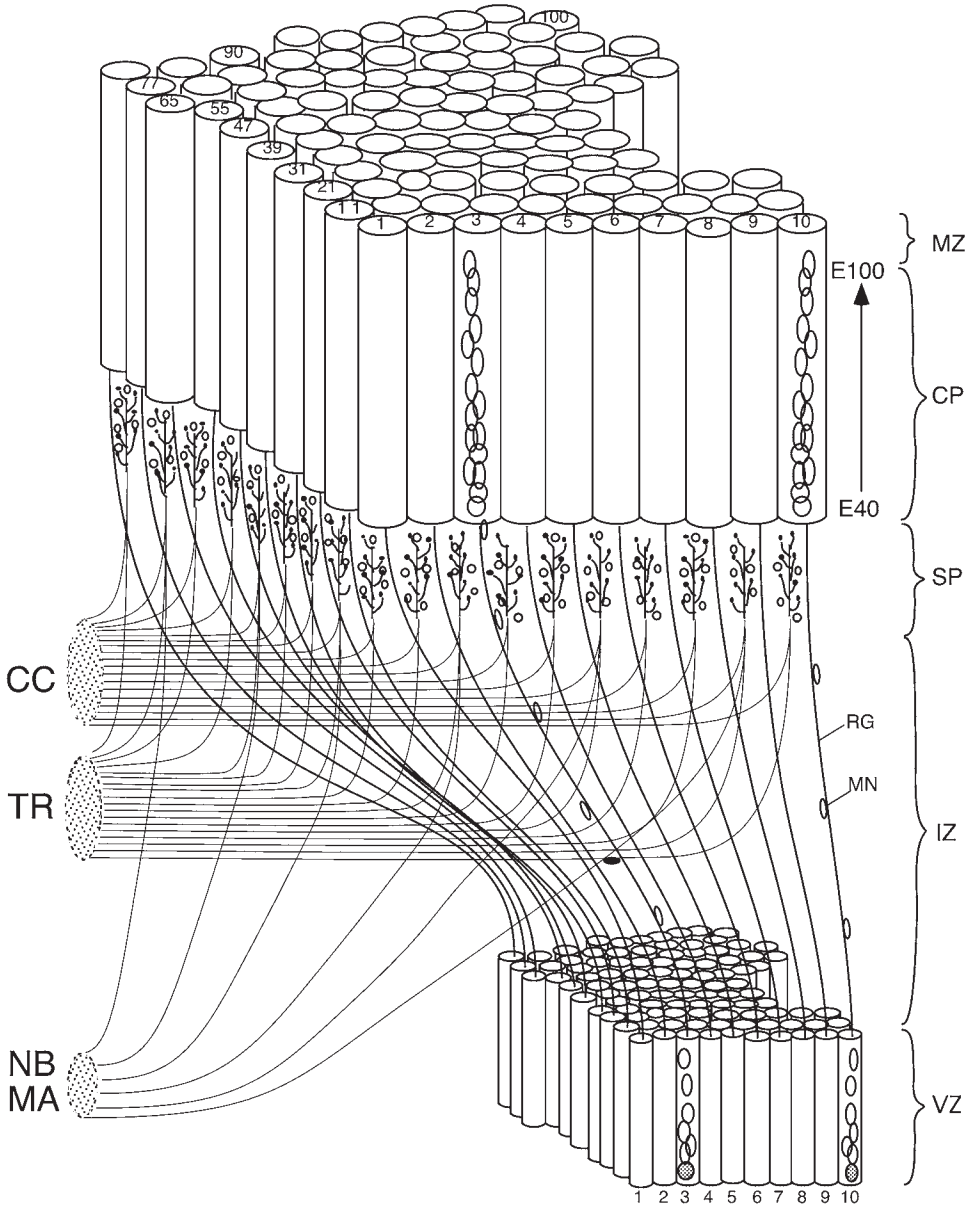


FIGURE 1.3 A three-dimensional illustration of the basic developmental events and types of cell-cell interactions occurring during the early stages of corticogenesis, before formation of the final pattern of cortical connections. This cartoon emphasizes radial migration, a predominant mode of neuronal movement which, in primates, underlies the elaborate columnar organization of the neocortex. After their last division, cohorts of migrating neurons (MN) first traverse the intermediate zone (IZ) and then the subplate zone (SP), where they have an opportunity to interact with "waiting" afferents arriving sequentially from the nucleus basalis and monoamine subcortical centers (NB, MA), from the thalamic radiation (TR), and from several ipsilateral and contralateral cortico-cortical bundles (CC). After the newly generated neurons bypass the earlier generated ones that are situated in the deep cortical layers, they settle at the interface between the developing cortical plate (CP) and the marginal zone (MZ), and eventually form a radial stack of cells that share a common site of origin but are generated at different times. For example, neurons produced between E40 and E100 in radial unit 3 follow the same radial glial fascicle and form ontogenetic column 3. Although some cells, presumably neurophilic in the nature of their surface affinities, may detach from the cohort and move laterally, guided by an axonal bundle (e.g., horizontally oriented, black cell leaving radial unit 3 and horizontally oriented fibers), most postmitotic cells are gliophilic; that is, they have an affinity for the glial surface and strictly obey constraints imposed by transient radial glial scaffolding (RG). This cellular arrangement preserves the relationships between the proliferative mosaic of the ventricular zone (VZ) and the corresponding protomap within the SP and CP, even though the cortical surface in primates shifts considerably during the massive cerebral growth encountered in midgestation. (For details see Rakic, 1988.)

lineage analysis of transcription factors specific for either dorsal or ventral neocortex progenitor cells suggested that the majority of interneurons in the human cortex are derived from the VZ/SVZ of the dorsal neocortex (Letinic et al., 2002; Petanjek et al., 2009; Petanjek et al., 2008). This finding strongly suggests that the primate cortex may confer species-specific differences in basic developmental events compared to other mammalian species.

Considerable progress has been made in understanding the molecular mechanisms behind neuronal migration and the physical displacement of the cell perikarya during somal translocation across the densely packed tissue. Initially, based on an observation in situ, it was proposed that a single pair of binding, complementary molecules with gliophilic properties can account for the recognition of glial guides (Rakic, 1981). However, in recent decades many classes of recognition and adhesion molecules have been discovered and are being tested (e.g., Cameron and Rakic, 1994; Anton, Cameron, & Rakic, 1996; Hatten & Mason, 1990; Schachner et al., 1985; reviewed in Hatten, 2002). The subject is too large to be reviewed in detail here. However, it is important to mention that voltage- and ligand-gated ion channels on the leading process and cell soma of migrating neurons regulate the influx of calcium ions into migrating neurons (Komuro & Rakic, 1992, 1993, 1996; Rakic & Komuro, 1995). Calcium fluctuations, in turn, may trigger polymerization of cytoskeletal and contractile proteins essential for cell motility and translocation of the nucleus and surrounding cytoplasm.

It was clear that translocation of nucleus requires cytoskeletal rearrangement (Rivas & Hatten, 1995; Rakic, Knyihar-Csillik, & Csillik, 1996), and a host of molecules involved in this complex process have been identified. For example, Doublecortin and Lis1 have been shown to be involved in cytoskeleton dynamics during neuronal migration (Gleeson, Lin, Flanagan, & Walsh, 1999; reviewed in Feng & Walsh, 2001), and their mutation in humans has been implicated in certain brain abnormalities such as Double Cortex (des Portes et al., 1998; Gleeson et al., 1998) and Lissencephaly Type I (Hattori, Adachi, Tsujimoto, Arai, & Inoue, 1994). Together these studies indicate that neuronal migration is a multifaceted developmental event, involving cell-to-cell recognition, differential adhesion, transmembrane signaling, and intracytoplasmic structural changes (Rakic et al., 1994).

Nuclear movement during both radial and tangential migration is a saltatory, two-step process alternating between resting and dynamic phases (Ang et al., 2003; Solecki, Model, Gaetz, Kapoor, & Hatten, 2004; Tsai & Gleeson, 2005). This process is evolutionarily conserved and involves a microtubule organizing center called centrosome that controls microtubule polymerization (Bornens, 2002). In radially migrating neurons, the centrosome controls the formation of a microtubule network that surrounds the nucleus—the so-called perinuclear cage—and establishes a physical link between centrioles and nuclear membrane (Rakic, 1971; Gregory, Edmondson, Hatten, & Mason, 1988; Rivas & Hatten, 1995; Rakic et al., 1996; Solecki et al., 2004; Higginbotham & Gleeson, 2007; Tsai, Bremner, & Vallee, 2007; reviewed in Baudoin, Alvarez, Gaspar & Metin, 2008). A similar mechanism has been described in cells of the subventricular zone that migrate tangentially to the olfactory bulb in the postnatal brain (Schaar & McConnell, 2005) and, more recently, in radially migrating cortical neurons (Tsai et al., 2007). Together, these results suggest that the actin-myosin cytoskeleton might control nuclear movements in neurons exhibiting multiple (radial and tangential) modes of migration (Baudoin et al., 2008).

A simple model of selective molecular components involved in cell migration is provided on the diagram in figure 1.4. The initial discovery of the glial-guided radial migration in primates led to the proposal of the *radial unit hypothesis* (Rakic, 1988), which has served as a useful working model for research on the cellular and molecular mechanisms involved in normal and abnormal cortical development and evolution and will be summarized in the next section.

Radial unit hypothesis

The radial unit hypothesis of cortical development postulates that the embryonic cortical plate forms from vertically oriented cohorts of neurons generated at the same site in the proliferative ventricular zone, called proliferative units, which consist of radially deployed progenitors in different phases of the cell cycle (Rakic, 1978). Thus each radial unit consists of several clones (polyclones) that produce postmitotic neurons, which migrate to the cortex following glial fascicles spanning the fetal cerebral wall and forming ontogenetic columns (Rakic, 1988). After arriving at the cortical plate, the later-generated cells bypass earlier-generated ones and settle in an inside-out gradient of neurogenesis (Rakic, 1974). Thus the two-dimensional positional information of the proliferative units in the ventricular zone is transformed into a three-dimensional cortical architecture: the *x*- and *y*-axis position of the cells is provided by their site of origin, whereas the *z*-axis position is provided by their time of origin (figure 1.3).

The radial unit hypothesis relies on the proposal that neurons comprising a given radial column are clonally related, and this idea could be tested experimentally with the aid of the retroviral gene transfer method for in vivo analysis of cell lineages in the mammalian brain (Sanes, 1989). Use of this approach suggested that most progenitors originating in the same site of the ventricular

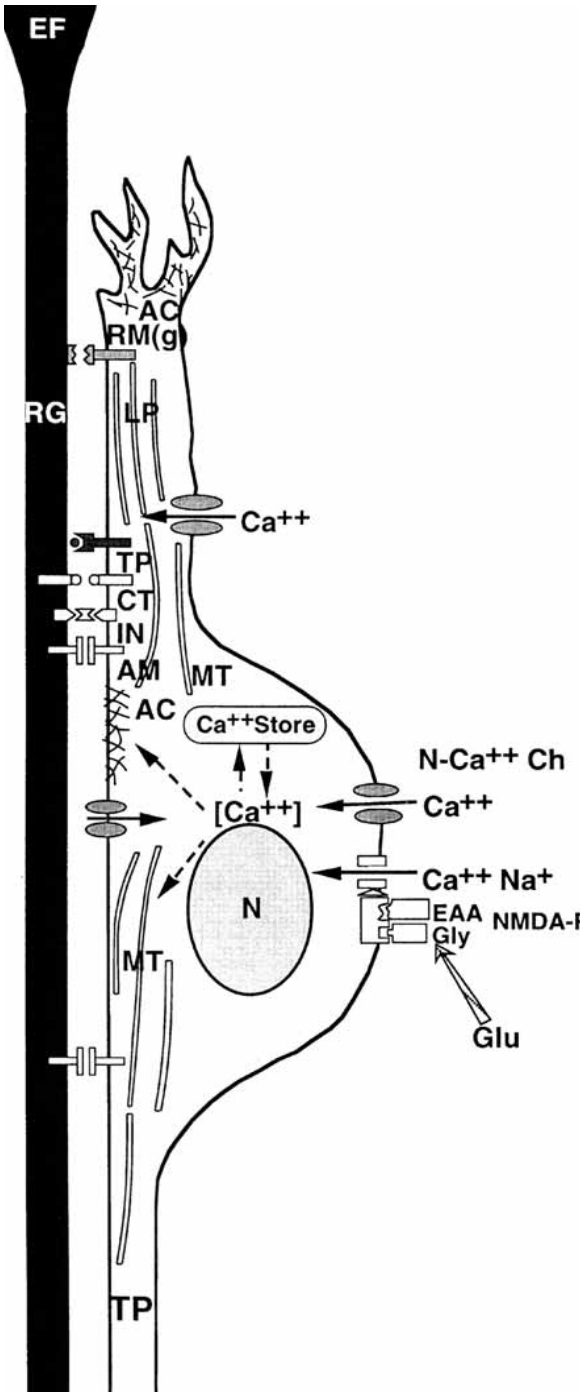


FIGURE 1.4 Model of a proposed cascade of cellular and molecular events that take place during the migration of postmitotic cells in the developing cerebral wall. After their last mitotic division in the ventricular zone, migrating cells extend a leading process (LP) that follows the contours of the radial glial fiber (RG) as it spans the expanding cerebral wall. The cytoskeleton within the LP and trailing process (TP) contains microtubules (MT) and actinlike contractile proteins (AC) that are involved in translocation of the cell nucleus (N) and the surrounding cytoplasm within the leading process until the cell enters the cortical plate. This system, maintained in vitro in slice preparations or imprint culture provides an opportunity to examine the role of the various molecules that are engaged in the recognition, adhesion, transmembrane signaling, and motility that underlie directed neuronal migration. The voltage-gated (N-type) and ligand-gated (NMDA-type) receptors/channels are thought to control calcium influx, which serve as messengers for execution of this movement. Abbreviations: AM, homotypic adhesion molecule; EAA, excitatory amino acid; EF, end foot of the radial glial fiber; Gly, glycine; LP, leading process; MT, microtubule; N, cell nucleus; TP, trailing process; RG, radial glial fiber; RM(g) gliophilic recognition molecule; TP, tyrosine phosphorylation. (Modified from Rakic, Cameron, & Komuro, 1994.)

zone remain radially deployed in the cortex (Luskin et al., 1988; Kornack & Rakic, 1995; Tan et al., 1998; see, however, Reid et al., 1995). Furthermore, a number of studies in chimeric and transgenic mice have provided evidence that a majority of postmitotic, clonally related neurons move and remain radially distributed in the cortex (e.g., Nakatsuji, Kadokawa, & Suemori, 1991; Soriano, Dumesnil, Auladell, Cohen-Tannoudji, & Sotelo, 1995; reviewed in Rakic, 1995a). The use of the retroviral gene transfer method in the embryonic primate brain showed that even in the large and highly convoluted cerebrum, radial deployment of many clones is remarkably well preserved (Kornack & Rakic, 1995).

The radial unit hypothesis provides a conceptual frame to understand how changes in the dynamics of proliferation, migration, and survival of newly generated cells in the proliferative regions translates into changes in the thickness or surface area of the cortex during individual development and evolution (Rakic, 1988, 1995b). The study of this developmental mechanism has been done mainly using transgenic mice and retroviral approaches in nonhuman primates (e.g., Kornack & Rakic, 1995; Kuida et al., 1996, 1998; Zhong, Feder, Jiang, Jan, & Jan, 1996; Haydar, Kuan, Flavell, & Rakic, 1999). Based on these data on the time of cell origin and cell proliferation kinetics, we proposed that the number of cortical cells and the size and morphology of the cortex depends on the mode of mitotic division (symmetric versus asymmetric), the duration of the cell cycle, and the degree of programmed cell death in the proliferative zones (Rakic, 1988, 1995b), as is described in the next section. Since most genes involved in cell production and fate determination seem to be preserved during evolution, one might expect that the control of neuronal number and differentiation would be basically similar in all species (R. Williams & Herrup, 1988).

Determinants of cortical size

The size of the cortex varies greatly among mammals, but it is enormously enlarged in primates. This growth has not been homogeneous and has consisted in a great increase in surface area and a little increase in thickness

(the surface area of the human cortex is a thousandfold larger than that of the mouse, but its thickness is only about two or three times that of the mouse cortex). This asymmetric growth seems to be related to the anatomical and functional structure of the neocortex that has been described as formed by iterative units (named radial columns, cortical columns, or cortical modules), consisting in a columnar array of neurons that share a common basic scheme of intrinsic and extrinsic connectivity and subserve the same function (Mountcastle, 1997; Szenthagothai, 1978; Goldman-Rakic, 1987). The larger the cortex in a given species, the larger the number of participating columnar units (Rakic, 1978, 1995b). The liberal use of the broad and undefined term "cortical column" has caused considerable confusion in the literature, since there is not a unique anatomical or functional columnar unit that applies to all cortical areas or species (Rakic, Ayoub, Dominguez, & Breunig, in press). In spite of this heterogeneity, there is consensus concerning the idea that the cortex is composed of an array of columnar units of operation.

Under this perspective, it seems that the addition of new functions to the cortex does not require a substantial increase in the thickness of the cortex, namely, in the number of components within the cortical units of operation (in spite of a very likely increase in their complexity), but would mainly correspond with the addition of new columnar units, which ultimately would contribute to the forming of new cortical regions and functional domains. In this respect, the radial unit hypothesis provides a mechanistic explanation of the large expansion of cortical surface without a concomitant significant increase in thickness during phylogenetic and ontogenetic cortical development (Rakic, 1988). It also shows how the genes controlling the number of founder cells in the mosaic of proliferative units at the ventricular surface set a limit on the size of the cortical surface during the development of the individual, as well as during the evolution of mammalian species (Rakic, 1995b). For example, a relatively small change in the timing of developmental cellular events could have large functional consequences: a minor increase in the length of the cell cycle or the magnitude of cell death in the ventricular zone could result in a large increase in the number of founder cells that form proliferative units (Rakic, 1988). Since proliferation in the ventricular zone initially proceeds exponentially because of the prevalence of symmetrical divisions, an additional round of mitotic cycles during this phase doubles the number of founder cells and, consequently, the number of ontogenetic radial columns (figure 1.5 and Rakic, 1995b). According to this model, fewer than four extra rounds of symmetrical cell divisions in the ventricular zone before the onset of corticogenesis can account for a tenfold difference in the cortical surface area. Since the mode of cell division changes to predominantly asymmetrical after the onset of corticogenesis, one can predict that the cell production period in humans will be about two weeks longer than in macaques, but should enlarge the cortical thickness by only 10% to 15%, a result which is actually observed (Rakic, 1995b). Thus, as illustrated in figure 1.5, even a small delay in the onset of the second phase of corticogenesis results in an order-of-magnitude larger cortical surface because of the increasing number of founder cells at the ventricular zone. This has been observed in mice expressing higher levels of β-catenin that enhance the number of founder cells and can lead to the formation of ectopias as well as an expanded cortical plate that begins to form convolutions (Chenn & Walsh, 2002, 2003).

In addition, the number of cortical neurons will depend on the degree of survival of the generated cells, and thus one mechanism that regulates the number of cells produced in the ventricular zone is programmed cell death (PCD) or apoptosis. Although PCD has been considered a major factor contributing to the formation of the vertebrate brain (Glucksmann, 1952), contemporary research has focused mainly on the histogenetic cell death involved in the elimination of inappropriate axonal connections at the later stages of development (e.g., Rakic & Riley, 1983a, 1983b; Oppenheim, 1991). However, the discovery of several classes of genes involved in apoptosis, which were initially identified in invertebrates, created the opportunity to study this phenomenon in the mammalian cerebrum. For example, a family of enzymes called caspases has been shown to play an important role in apoptosis in a variety of organs and tissues (Ellis & Horvitz, 1991). We have demonstrated that in mouse embryos deficient in caspase 9 and 3, fewer cells are eliminated than in their littermates (Kuida et al., 1996, 1998). Reduction of apoptosis in the knockout mice for these caspases results in the formation of supernumerary founder cells in the cerebral ventricular zone. As a consequence, these mice form ectopic cells in the intermediate zone as well as a larger cortical plate with more radial units. Correspondingly, diminishment of the clearance rate of apoptotic cells in transgenic mice lacking a receptor crucial for the recognition of dying cells results in enlarged proliferative zones that lead to the formation of ectopias as well as an expanded cortical plate that begins to form convolutions (Haydar et al., 1999).

In both cases, by increase of the founder population or by reduction of apoptosis, the result is a larger cortical sheet that begins to buckle and forms incipient convolutions, without significant change in thickness of the developing cortical plate (Haydar et al., 1999; Chenn & Walsh, 2002). This result is a good example of how the mutation of a few key genes that control the reduction of cells could result in the expansion of the cortex and the appearance of convolutions during the evolution of the cerebral cortex (Rakic, 1995b).

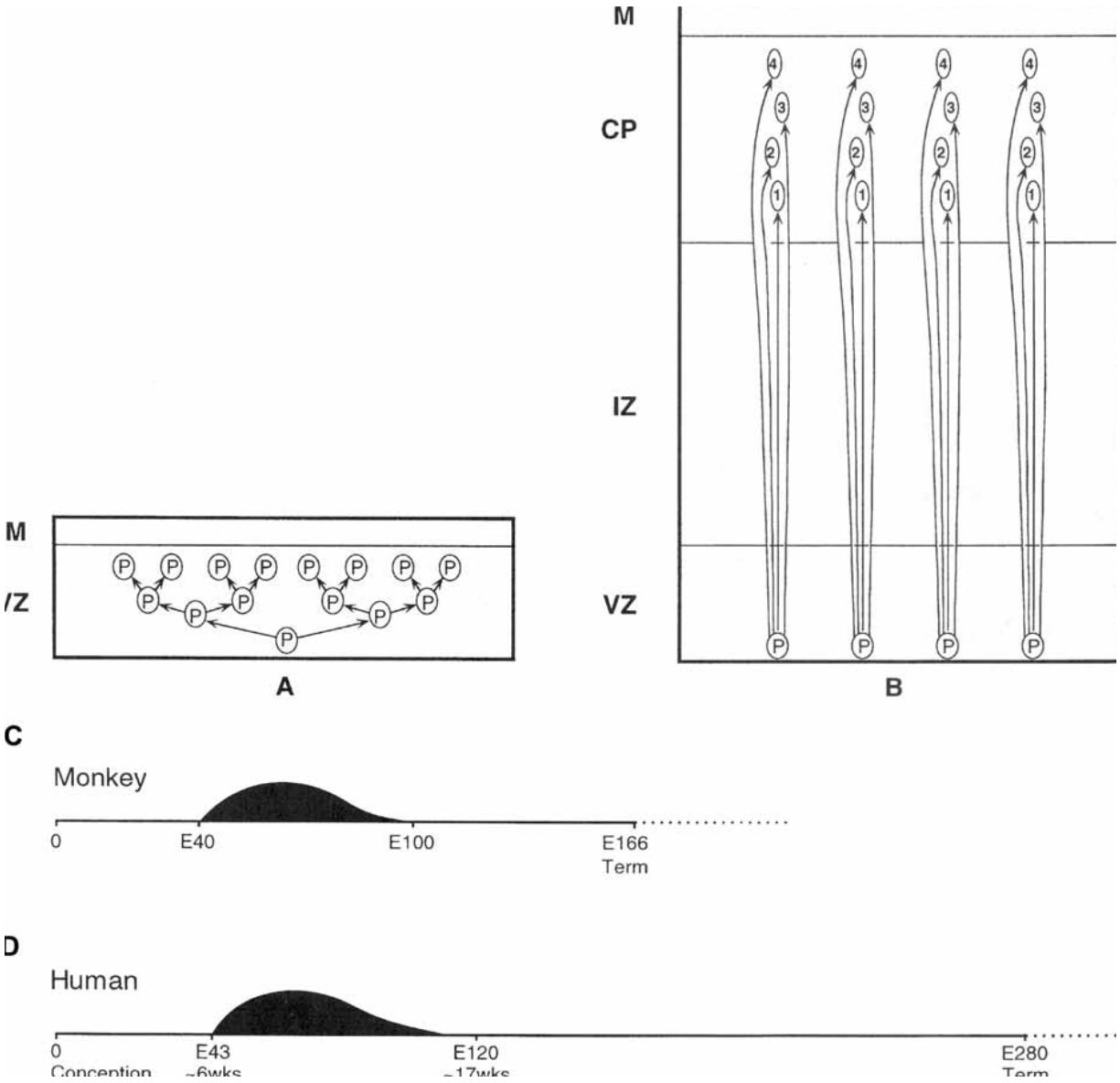


FIGURE 1.5 (*A*) Schematic model of symmetrical cell divisions that predominate before E40. At this early embryonic age, the cerebral wall consists of only the ventricular zone (VZ), where all the cells are proliferating, and the marginal zone (M), into which some cells extend radial processes. Symmetric division produces two progenitors (P) during each cycle and causes rapid horizontal lateral spread. (*B*) Model of asymmetrical or stem division that becomes predominant in the monkey embryo after E40. During each asymmetrical division a progenitor (P) produces one postmitotic neuron that leaves the ventricular zone and another progenitor that remains within the proliferative zone and continues to divide. Postmitotic neurons migrate rapidly across the intermediate zone (IZ) and become arranged vertically in the cortical plate (CP) in reverse order of their arrival (1, 2, 3, and 4). (*C*) Diagrammatic representation of the time of neuron origin in the macaque monkey. The data were obtained from ^{3}H-thymidine autoradiographic analyses (from Rakic, 1995b). (*D*) Estimate of the time of neuron origin in the human neocortex based on the number of mitotic figures within the ventricular zone, supravital DNA synthesis in slice preparations of fetal tissue, and the presence of migrating neurons in the intermediate zone of the human fetal cerebrum. (From Rakic, 1995b.)

Protomap hypothesis

A major challenge to students of the cerebral cortex is how individual and species-specific cytoarchitectonic areas have emerged from the initially seemingly uniform ventricular zone and cortical plate. Both intrinsic and extrinsic factors have been suggested. One attractive model, known as the *tabula rasa* hypothesis, is that all cortical neurons are equipotent and that laminar and areal differences are induced by extrinsic influences exerted via thalamic afferents (Creutzfeldt, 1977). However, there was also considerable evidence that the cells generated within the embryonic cerebral wall contained some intrinsic information about their prospective species-specific cortical organization. The *protomap hypothesis* was proposed as an empirical generalization in an attempt to reconcile available experimental and descriptive data in this field (Rakic, 1988). It was based on the fact that all ontogenetic columns are mixtures of several clones and are not equal in each prospective cytoarchitectonic area (footnote 13 in Rakic, 1988). The protomap model suggests that the basic pattern of cytoarchitectonic areas emerges through synergistic, interdependent interactions between developmental programs intrinsic to cortical neurons and extrinsic signals supplied by specific inputs from subcortical structures. According to this hypothesis, neurons in the embryonic cortical plate—indeed even in the proliferative ventricular zone where they originate—set up a primordial map that preferentially attracts appropriate afferents and has a capacity to respond to this input in a specific manner. The prefix *proto* was introduced to emphasize the primordial, provisionary, and essentially malleable character of the *map*, which is subject to considerable modification by the extrinsic influences exerted at later stages (Rakic, 1988).

The initial indication that developmental events in the proliferative ventricular zone foreshadow prospective regional differences in the overlying cerebral mantle comes from the observation that the neurogenesis of the primary visual cortex, which contains more neurons per radial unit than the adjacent areas, lasts longer (Rakic, 1976a). Furthermore, it has also been demonstrated that the mitotic index in the ventricular region subjacent to this area is higher than in adjacent regions (Dehay, Giroud, Berland, Smart, & Kennedy, 1993). In addition, initial establishment of cytoarchitectural and functional features specific to the visual cortex, such as the ocular dominance columns, form independently of thalamic inputs to this area (Crowley & Katz, 2000). Therefore, certain region-specific differences in the production of the ventricular zone can be detected even before neurons arrive at the cortex (Rakic, 1988; Kennedy & Dehay, 1993; Algan & Rakic, 1997). In addition, several lines of evidence indicate that, during the final cell division, one or both daughter cells start to express a variety of neuron-class-specific signaling molecules (LoTurco, Owens, Heath, Davis, & Kriegstein, 1995; Lidow & Rakic, 1994). Postmitotic cells not only become committed to a neuronal fate but also become restricted in their repertoire of possible fates (McConnell, 1988). Numerous studies in which the cytology of postmitotic cells has been examined (e.g., Schwartz, Rakic, & Goldman-Rakic, 1991; LoTurco et al., 1995) and/or manipulated by a variety of methods such as spontaneous mutations (e.g., Caviness & Rakic, 1978: Rakic, 1995b), ionizing radiation (Algan & Rakic, 1997), retroviral gene transfer labeling (Parnavelas, Barfield, Franke, & Luskin, 1991), transgenic inserts (Kuida et al., 1996), and heterochronic transplantations (McConnell, 1988; McConnell & Kaznowski, 1991) all indicate that certain class-specific cell attributes are expressed before the migrating neurons arrive at the cortical plate and become synaptically connected. Remarkably, neurotransmitter secretion is unnecessary for proper brain formation (Verhage et al., 2000). In addition, retroviral tracing experiments and some clonal analyses suggest that the ventricular zone is comprised of a heterogeneous population of cells, and that cell lineage contributes substantially to the cell fate determination of neurons (Acklin & van der Kooy, 1993; Parnavelas et al., 1991; Kornack & Rakic, 1995; B. Williams & Price, 1995; Kuan, Elliott, Flavell, & Rakic, 1997). Emerging observations confirm and extend these findings both within the cortex, in migrating interneurons, and later postnatally (Rakic et al., in press; Merkle, Mirzadeh, & Alvarez-Buylla, 2007; Batista-Brito, Machold, Klein, & Fishell, 2008).

These findings raise the question of whether laminar and areal identities of cortical plate cells provide cues or chemotactic attractants for incoming afferent axons. Data from axonal tracing indicate that afferent connections from subcortical structures and other cortical regions find their way to the specific regions of the cortical plate either directly or by way of the subplate zone or both (Kostovic & Rakic, 1984, 1990; De Carlos & O'Leary, 1992; McConnell et al., 1994; Agmon, Yang, Jones, & O'Dowd, 1995; Catalano, Robertson, & Killackey, 1996; Richards, Koester, Tuttle, & O'Leary, 1997), suggesting the existence of region-specific attractants for pathfinding and target recognition. In support of this idea the development of correct topological connections in anophthalmic mice and in early enucleated animals indicates that basic connections and chemoarchitectonic characteristics can form in the absence of information from the periphery (e.g., Kaiserman-Abramof, Graybiel, & Nauta, 1980; Olavarria & van Sluyters, 1984; Rakic, 1988; Kennedy & DeHay, 1988; Kuljis & Rakic, 1990; Rakic & Lidow, 1995; Miyashita-Lin, Hevner, Wassarman, Martinez, & Rubenstein, 1999).

The gradients or region-specific distribution of various morphoregulatory molecules (or both) in the embryonic cerebral wall (e.g., Arimatsu et al., 1992; Levitt, Barbe, & Eagleson, 1997; Ferri & Levitt, 1993; Porteus, Bulfone,

Ciaranello, & Rubenstein, 1991; Bulfone et al., 1993; Cohen-Tannoudji, Babinet, & Wassef, 1994; Emerling & Lander, 1994; Donoghue & Rakic, 1999; Sestan, Rakic, & Donoghue, 2001; Bishop, Goudreau, & O'Leary, 2000) or layer-specific expression of POU-homeodomain genes (e.g., Frantz, Bohner, Akers, & McConnell, 1994; Meissirel, Wikler, Chalupa, & Rakic, 1997) may also contribute to the formation of specified axonal pathways. For example, electroporation of FGF8 applied before beginning of formation of thalamic connections has shown that morphoregulatory molecules can shift the anterior/posterior areal boundaries in the developing neocortex (Fukuchi-Shimogori & Grove, 2001). Thus tangentially and radially distinct landmarks in the postmitotic cells facilitate axonal pathfinding and target recognition that eventually lead to parcellation of the cerebral cortex.

It should be underscored that although the embryonic cerebral wall exhibits gradients of several morphoregulatory molecules, as well as other specific area-specific molecular differences, the protomap within the embryonic cerebrum provides only a set of species-specific genetic instructions and biological constraints. The precise position of interareal borders, the overall size of each cytoarchitectonic area, and the details of their cellular and synaptic characteristics in the adult cerebral cortex are achieved through a cascade of reciprocal interactions between cortical neurons and the cues they receive from afferents arriving from a variety of extracortical sources (Rakic, 1988). Such afferents may serve to coordinate and adjust the ratio of various cell classes with the subcortical structures, as has been shown in the primary visual system (Meissirel et al., 1997; Rakic, Suner, & Williams, 1991; Kennedy & Dehay, 1993; Rakic & Lidow, 1995). In summary, the concept of the cortical protomap includes the role of both intrinsic and extrinsic determinants in shaping the final pattern and relative sizes of the cytoarchitectonic areas.

Initial formation of synaptic connections

Acquiring areal and laminar positions and phenotypes comprises only the first and most fundamental steps in the formation of species-specific cortical development. However, an equally important next step is the formation of synaptic connections. It is a very large subject, is studied by most developmental neurobiologists, and by itself deserves more than a short section. It is in part covered in the subsequent chapters, and it is only briefly reviewed here to emphasize that in primates, and particularly in the human, it is a prolonged process that involves overproduction of cells, axons, and synapses and their later elimination in response to environmental influences. This process is often called neuroplasticity, and we consider it as a phase of the normal course of brain development. In the macaque monkey it peaks after birth and lasts at least 3–4 years, and in the human it lasts 15–19 years.

Detailed studies in the rhesus monkey showed that both neurons and their axons are overproduced in the cerebral cortex during well-delineated stages of development. For example, there are about 40% more neurons in the monkey visual cortex during the second half of pregnancy than in the adult. Furthermore, a newborn monkey has almost 200 million callosal axons compared to less than 50 million in the adult (LaMantia & Rakic, 1990). The axons are lost at the rate of about 8 million per day or 50 per second during the first three weeks after birth. Thereafter, they are lost at an estimated rate of half a million per day or 5 per second until the adult value is reached. Similar magnitude of elimination of callosal axons has been reported for other mammalian species (e.g., Berbel & Innocenti, 1988). The functional significance of this loss of axons is not fully understood, but the prevailing hypothesis has been that activity-dependent stabilization plays a critical role.

In the rhesus monkey, during the first two to three months of postnatal life, synaptic density increases rapidly and reaches a peak that is about two times higher than in the adult. This synaptic density well above the adult level lasts throughout infancy and adolescence. We have calculated that about 1.8×10^{11} synapses are lost in the visual cortex of a single cerebral hemisphere during puberty in the monkey (Bourgeois & Rakic, 1993). The magnitude of this decline is stunning when expressed as loss of about 2500 synapses per second during this period. Since other areas, including association cortices, simultaneously undergo comparable synaptic loss, more than 30,000 synapses per second are deleted from the entire cortical mantle during monkey adolescence (Bourgeois & Rakic, 1993). In the human, in whom this period of life last about three times longer but the cortex is ten times larger, the synapses may be lost at the rate of 100,000 per second.

The decline in synaptic density is due primarily to elimination of excitatory junctions located on dendritic spines, while inhibitory synapses on dendritic shafts remain relatively constant. Other results showed that the density of major neurotransmitter receptors in the cortex also reaches a maximum level between two and four months of age and then declines to the adult level during the period of sexual maturation (Lidow, Goldman-Rakic, & Rakic, 1991). These findings revealed an unusual coordination between biochemical and structural differentiation and indicated that these events may be related to maturation of function.

The prolonged phase of postnatal development in primates provides unparalleled opportunity for competitive activity-driven stabilization among initially supernumerary inter- and intracortical connections. The formation of the final pattern of cortical connections is achieved through dynamic interactions that involve at least two well-defined

steps. In the first step, one set of axons project to the target structure guided by prespecified short-lived molecular markers, without regard to specific location on the individual neurons or their parts (Easter, Purves, Rakic, & Spitzer, 1985). In the second step, which is activity dependent, synaptic connections are sorted out and remain on only selective sets of neurons or their dendrites. Our experimental studies of the formation of binocular vision in primates provide a dramatic example of this biphasic development and support the hypothesis that competitive interactions between two or more populations of neurons play a significant role in the elimination of the axons and segregation of their synapses (Rakic, 1976a, 1977, 1981).

This phase of cortical development has a profound significance for understanding development of human cognitive abilities. Some implications are obvious. For example, our educational system, which postpones disciplined intellectual learning until later stages in life, may not be biologically optimal, as the first 15 years may be the most important formative phase as far as synaptic stabilization is concerned. It is well known that high-level professional musicians or athletes can be created only at a specific critical age, and there is no reason why other cognitive skills essential for human intellectual abilities are not set up also at this time. Our data provide biological support for intensive targeted intellectual training that starts in childhood if we expect optimal results.

Timing of cortical genesis in the human

Since our ultimate goal is understanding the development of the neocortex in humans, the comparison of selected cellular features of the human cortex at different prenatal stages with that of the macaque monkey may help to determine the corresponding time and sequence of developmental events in these species. This determination is essential if we want to apply the findings obtained from experimental animals to the understanding of human cortical development (e.g., Chalupa & Wefers, 2000). To this end, Poliakov's comprehensive histological studies of cortical development in human fetuses, published originally in the Russian literature (e.g., Poliakov, 1959, 1965) have been reviewed in more detail elsewhere (Sidman & Rakic, 1982), but are presented here summarized in figure 1.6 and compared with the timing of corresponding events in the macaque monkey.

Stage I Initial formation of the cortical plate (from approximately the 6th to the 10th fetal weeks). During the 7th fetal week, postmitotic cells begin to migrate from the ventricular zone outward to form a new accumulation of cells at the junction of the intermediate and marginal zones. By the middle of this period, synapses of unknown origin are present above and below the cortical plate (Molliver, Kostovic, & van der Loos, 1973; Zecevic, 1998). This stage corresponds approximately to the level of cortical development found in the monkey fetus between E40 and E54, depending on the region.

Stage II Primary condensation of the cortical plate (through approximately the 10th and 11th fetal weeks). At this stage the cortical plate increases in thickness, becomes more compact, and is clearly demarcated from the fiber-rich part of the intermediate zone, which seems to have fewer cells per unit volume, indicating that the first major wave of migration is almost spent (figure 1.6). The end of this stage corresponds approximately to the E55–E59 period in the monkey when the majority of neurons of layers 5 and 6 are generated in most regions of the cortex (Sidman & Rakic, 1982; Marin-Padilla, 1988).

Stage III Bilaminate cortical plate (most pronounced during the 11th to the 13th fetal week). The uniform and compact cortical plate of the second stage becomes subdivided into an inner zone occupied mainly by cells with relatively large, somewhat widely spaced oval nuclei and an outer zone of cells with densely packed, darker, bipolar nuclei (figure 1.6). This heterogeneity results from the more advanced maturation of the deep-lying neurons that had arrived at the cortical plate during earlier developmental stages, plus the addition of a new wave of somas of immature neurons that take up more superficial positions. This period is also characterized by the appearance of the cell-sparse, fiber-rich subplate zone situated below the cortical plate. This transient embryonic zone in the human fetus is particularly wide in the regions subjacent to the association areas (Kostovic & Rakic, 1990). The third stage corresponds roughly to the level of development achieved in the monkey between E59 and E64.

Stage IV Secondary condensation (from the 13th to the 15th fetal week). During this period of gestation, the ventricular zone becomes progressively thinner, while the subventricular zone remains relatively wide (figure 1.6). The cortical plate again becomes homogeneous in appearance and resembles, in a sense, a thickened version of Stage II. The reason for this change may be that, in Stage IV, most of the young neurons in the cortex become considerably larger as they differentiate, while relatively few new immature neurons enter the cortical plate. The result is a more uniform appearance. At the end of this stage, an accumulation of large cells appears below the cortical plate, and the subplate zone enlarges further (Kostovic & Rakic, 1990). Depending on the cortical region, this stage appears in the monkey between E64 and E75.

Stage V Prolonged stage of cortical maturation (from the 16th fetal week continuing well into the postnatal period).

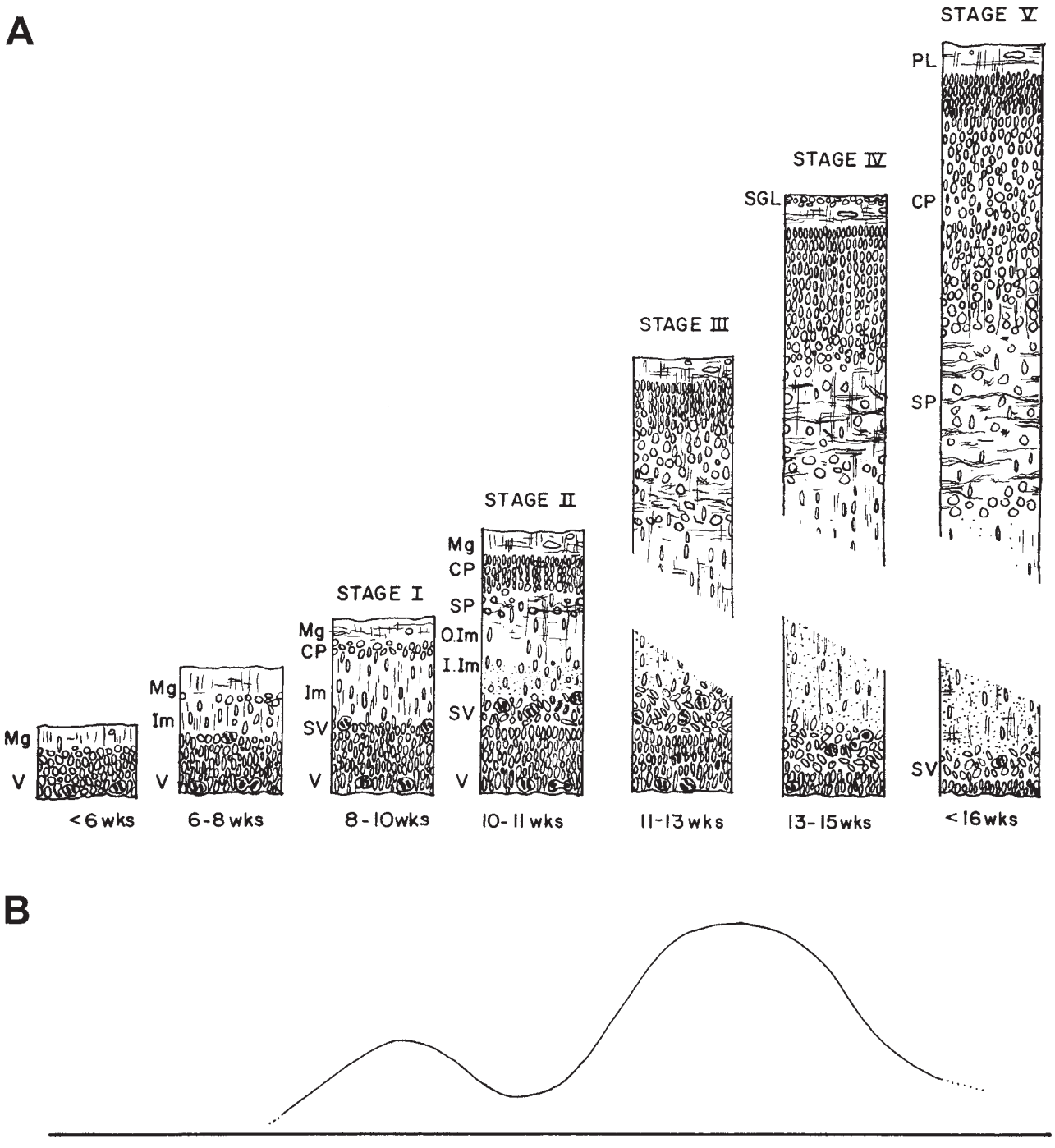


FIGURE 1.6 (*A*) Semidiagrammatic drawings of the human cerebral wall at various gestational ages listed in fetal weeks below each column. The stages refer specifically to an arbitrarily chosen cortical area situated midway along the lateral surface of the hemisphere (detailed in Sidman & Rakic, 1982). Because there is a gradient of maturation, as many as three of five stages of cortical development may be observed in different regions of the neocortex in the same fetal brain. In the three columns on the right, the intermediate zone is not drawn in full because the thickness of the cerebral wall has increased markedly compared with earlier stages and cannot fit into the drawing. In addition, in the last three stages the subplate zone, situated below the cortical plate, appears (Kostovic & Rakic, 1990). (*B*) The curve below the drawing schematically indicates waves of cell migration to the neocortex assessed by the density of migrating neurons in the intermediate zone. Abbreviations: CP, cortical plate; Im, intermediate zone; I.Im and O.Im, inner and outer intermediate zones, respectively; Mg, marginal zone; PL, plexiform layer; SGL, subpial granular layer; SP, subplate zone; SV, subventricular zone; V, ventricular zone; wks, age in fetal weeks. (From Rakic, 1988.)

Morphological data are inadequate to determine for how long neurons continue to migrate to the human neocortex after 16 weeks or how many do so, and hence the line at the right side of the curve is dotted in figure 1.6*B*. By the fifth month, relatively few neuronal precursors seem to be proliferating in the reduced ventricular zone of the human cerebral hemispheres. However, the interneurons, which continue to be generated in the subventricular zone and ganglionic eminence, are still being added to the cortex between the 20th and 25th weeks of gestation (Letinic et al., 2002). A comparison of the autoradiographic results in the monkey (Rakic, 1974, 1977) with comparable stages in human (Rakic & Sidman, 1968; Marin-Padilla, 1988; Kostovic & Rakic, 1990) indicates that most neurons of the human neocortex are generated before the beginning of the third trimester of gestation. Toward term, the ventricular zone disappears, the subplate zone dissolves, and as the intermediate zone transforms into the white matter, only a vestige of the subplate cells remain as interstitial neurons (Kostovic & Rakic, 1980). It should be emphasized that during neurogenesis and after completion of neurogenesis, glial cells are generated and migrate across the cerebral wall. This phenomenon is particularly pronounced in the human forebrain where glial cells including oligos largely outnumber neurons. After all cortical neurons have been generated and attained their final positions, their differentiation, including the formation of synapses, proceeds for a long time and reaches a peak only during the second postnatal year. The subject of synaptogenesis in the cerebral cortex of both macaque monkeys and humans was reviewed in the second edition of this volume (Bourgeois, Goldman-Rakic, & Rakic, 2000) and not much new research has been done in this area since.

The human factor

As mentioned before, traditional research on the cerebral cortex of mammals has recognized the quantitative differences (e.g., in the period of neurogenesis, number of progenitors, and duration of the cell cycle) that simply relate to the expansion of the forebrain, and particularly the great increase in size of cerebral cortex in primates, although other unique human characteristics have been neglected, perhaps because so few investigators work on the development of the human cerebrum. In this chapter we have exposed in some detail those differences, giving some insight on their possible contribution to the specification of the human cortical features. To finish this chapter, we would like to summarize those differences—and some others not mentioned before—that we think might contribute to establishing the differences in the cognitive potential of the primate, and specifically of the human, cortex.

• There is a pronounced gradient of maturation across the large hemispheres in primates (Donoghue & Rakic, 1999) that is much less visible in species with smaller forebrains, and especially in species with marked altricial growth strategies, such as mice and rats, that exhibit a short and fast brain development.

• Radial glial cells are particularly prominent in the embryonic primate cerebrum; they start expressing GFAP at the very start of corticogenesis (Levitt & Rakic, 1980; Levitt et al., 1981), and a subset of them transiently stop dividing (Schmechel & Rakic, 1979).

• The primate SVZ is dramatically enlarged and exhibits much more complexity in cellular organization in primates than in other species (Smart et al., 2002; Bystron et al., 2008). Another related feature is the appearance, in midgestation in human (~22 weeks) and monkey (~E72), of a band of tangentially orientated axons that divides the SVZ into inner and outer sublayers in some parts of the cortex (Bystron et al., 2008).

• A subpial granular layer has been described in humans that has not been observed in rodents. This is a long-known feature of human development (e.g., Brun, 1965) that is however systematically ignored at the present time.

• The reported differences in the developmental processes between species are reflected in the final cellular output, and, for example, the proportion of GABAergic interneurons exhibits a marked difference between rodents and primates: GABAergic interneurons represent approximately 15% of the total population of neurons in the rat neocortex (Beaulieu, 1993; Micheva & Beaulieu, 1995; Meinecke & Peters, 1987), whereas in the primates and the human, this proportion reaches 20% in the visual cortex and up to 25% in other cortical areas (Hendry, Schwark, Jones, & Yan, 1987; Beaulieu, Kisvarday, Somogyi, Cynader, & Cowey, 1992; Del Rio & DeFelipe, 1996). In addition, some GABAergic cell types like the double-bouquet cells are absent in rodents but can be detected in carnivores and are much more abundant and well developed in primates and more specifically in humans (Yanez et al., 2005). Double-bouquet interneurons are characterized by the presence of a long and narrow descending bundle of axons that are distributed following a microcolumnar pattern in the neocortex. The axons of double-bouquet cells establish hundreds of synapses with a narrow column of tissue, and therefore they have been proposed as key elements in the microcolumnar organization of the cortex, acting on pyramidal cells from different layers within the minicolumns (DeFelipe, Hendry, Hashikawa, Molinari, & Jones, 1990; Jones, 2000; Yanez et al., 2005).

Regarding excitatory neurons, differences have been reported in the size and pattern of arborization of projection neurons: pyramidal cells from the human prefrontal cortex

are bigger than those from other species, have larger and more complex basal dendritic fields, and also have a higher density of dendritic spines (Campbell & Morrison, 1989; Benavides-Piccione, Ballesteros-Yanez, DeFelipe, & Yuste, 2002; Elston, Benavides-Piccione, & DeFelipe, 2001; Elston et al., 2006). This increase is beyond scaling of cell structure and reflects specific specializations that allow human pyramidal cells from highly associative cortices to receive more synaptic inputs from more diverse origins, and therefore to be capable of integrating more information (Elston, 2003). In addition, specific subtypes have been described, such as the spindle cells that seem to be specific to some mammalian orders, and these are particularly abundant in restricted cortical regions of great apes and especially in humans (Nimchimsky et al., 1999; Allman, Hakeem, & Watson, 2002; Marino et al., 2007).

These marked differences in the cellular composition of the neocortex, from predecessor cells to GABAergic interneurons or glial cells, have to be taken into consideration if we are poised to understand the distinctively human aspects of behavior and cognition and the uniquely human cognitive disorders such as autism, dyslexia, or schizophrenia that probably involve genetic anomalies affecting the regulation of human-specific developmental features.

ACKNOWLEDGMENTS This work was supported by the U.S. Public Health Service.

REFERENCES

Acklin, S. E., & van der Kooy, D. (1993). Clonal heterogeneity in the germinal zone of the developing rat telencephalon. *Development*, *118*, 175–192.

Agmon, A., Yang, L. T., Jones, E. G., & O'Dowd, D. K. (1995). Topological precision in the thalamic projection to neonatal mouse barrel cortex. *J. Neurosci.*, *15*, 549–561.

Algan, O., & Rakic, P. (1997). Radiation-induced, lamina-specific deletion of neurons in the primate visual cortex. *J. Comp. Neurol.*, *381*, 335–352.

Allman, J., Hakeem, A., & Watson, K. (2002). Two phylogenetic specializations in the human brain. *Neuroscientist*, *8*, 335–346.

Altman, J., & Bayer, S. A. (1990). Vertical compartmentation and cellular transformations in the germinal matrices of the embryonic rat cerebral cortex. *Exp. Neurol.*, *107*, 23–35.

Anderson, S. A., Kaznowski, C. E., Horn, C., Rubenstein, J. L., & McConnell, S. K. (2002). Distinct origins of neocortical projection neurons and interneurons in vivo. *Cereb. Cortex*, *12*, 702–709.

Ang, E. S., Jr., Gluncic, V., Duque, A., Schafer, M. E., & Rakic, P. (2006). Prenatal exposure to ultrasound waves impacts neuronal migration in mice. *Proc. Natl. Acad. Sci. USA*, *103*, 12903–12910.

Ang, E. S., Jr., Haydar, T. F., Gluncic, V., & Rakic, P. (2003). Four-dimensional migratory coordinates of GABAergic interneurons in the developing mouse cortex. *J. Neurosci.*, *23*, 5805–5815.

Angevine, J. B., Jr., & Sidman, R. L. (1961). Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. *Nature*, *192*, 766–768.

Anton, E. S., Cameron, R. S., & Rakic, P. (1996). Role of neuron-glial junctional domain proteins in the maintenance and termination of neuronal migration across the embryonic cerebral wall. *J. Neurosci.*, *16*, 2283–2293.

Arimatsu, Y., Miyamoto, M., Nihonmatsu, I., Hirata, K., Uratani, Y., Hatanaka, Y., et al. (1992). Early regional specification for a molecular neuronal phenotype in the rat neocortex. *Proc. Natl. Acad. Sci. USA*, *89*, 8879–8883.

Batista-Brito, R., Machold, R., Klein, C., & Fishell, G. (2008). Gene expression in cortical interneuron precursors is prescient of their mature function. *Cereb. Cortex*, *18*, 2306–2317.

Baudoin, J. P., Alvarez, C., Gaspar, P., & Metin, C. (2008). Nocodazole-induced changes in microtubule dynamics impair the morphology and directionality of migrating medial ganglionic eminence cells. *Dev. Neurosci.*, *30*, 132–143.

Beaulieu, C. (1993). Numerical data on neocortical neurons in adult rat, with special reference to the GABA population. *Brain Res.*, *609*, 284–292.

Beaulieu, C., Kisvarday, Z., Somogyi, P., Cynader, M., & Cowey, A. (1992). Quantitative distribution of GABA-immunopositive and -immunonegative neurons and synapses in the monkey striate cortex (area 17). *Cereb. Cortex*, *2*, 295–309.

Benavides-Piccione, R., Ballesteros-Yanez, I., DeFelipe, J., & Yuste, R. (2002). Cortical area and species differences in dendritic spine morphology. *J. Neurocytol.*, *31*, 337–346.

Berbel, P., & Innocenti, G. M. (1988). The development of the corpus callosum in cats: A light- and electron-microscopic study. *J. Comp. Neurol.*, *276*, 132–156.

Bhardwaj, R. D., Curtis, M. A., Spalding, K. L., Buchholz, B. A., Fink, D., Bjork-Eriksson, T., et al. (2006). Neocortical neurogenesis in humans is restricted to development. *Proc. Natl. Acad. Sci. USA*, *103*, 12564–12568.

Bielle, F., Griveau, A., Narboux-Neme, N., Vigneau, S., Sigrist, M., Arber, S., et al. (2005). Multiple origins of Cajal-Retzius cells at the borders of the developing pallium. *Nat. Neurosci.*, *8*, 1002–1012.

Bishop, K. M., Goudreau, G., & O'Leary, D. D. (2000). Regulation of area identity in the mammalian neocortex by Emx2 and Pax6. *Science*, *288*, 344–349.

Bornens, M. (2002). Centrosome composition and microtubule anchoring mechanisms. *Curr. Opin. Cell Biol.*, *14*, 25–34.

Boulder Committee, American Association of Anatomists (1970). Embryonic vertebrate central nervous system: Revised terminology. *Anat. Rec.*, *166*, 257–261.

Bourgeois, J. P., Goldman-Rakic, P. S., & Rakic, P. (2000). Formation, elimination and stabilization of synapses in the primate cerebral cortex. In M. S. Gazzaniga (Ed.), *Cognitive neuroscience: A handbook for the field* (pp. 23–32). Cambridge, MA: MIT Press.

Bourgeois, J. P., & Rakic, P. (1993). Changes of synaptic density in the primary visual cortex of the macaque monkey from fetal to adult stage. *J. Neurosci.*, *13*, 2801–2820.

Breunig, J. J., Arellano, J. I., Macklis, J. D., & Rakic, P. (2007). Everything that glitters isn't gold: A critical review of postnatal neural precursor analyses. *Cell Stem Cell*, *1*, 612–627.

Breunig, J. J., Silbereis, J., Vaccarino, F. M., Sestan, N., & Rakic, P. (2007). Notch regulates cell fate and dendrite morphology of newborn neurons in the postnatal dentate gyrus. *Proc. Natl. Acad. Sci. USA*, *104*, 20558–20563.

Brun, A. (1965). The subpial granular layer of the foetal cerebral cortex in man. *Acta Pathol. Microbiol. Scand., Suppl. 179*, 173–198.

Bulfone, A., Kim, H. J., Puelles, L., Porteus, M. H., Grippo, J. F., & Rubenstein, J. L. (1993). The mouse Dlx-2 (Tes-1) gene is expressed in spatially restricted domains of the forebrain, face and limbs in midgestation mouse embryos. *Mech. Dev., 40*, 129–140.

Bystron, I., Blakemore, C., & Rakic, P. (2008). Development of the human cerebral cortex: Boulder Committee revisited. *Nat. Rev. Neurosci., 9*, 110–122.

Bystron, I., Rakic, P., Molnar, Z., & Blakemore, C. (2006). The first neurons of the human cerebral cortex. *Nat. Neurosci., 9*, 880–886.

Cameron, R. S., & Rakic, P. (1991). Glial cell lineage in the cerebral cortex: A review and synthesis. *Glia, 4*, 124–137.

Cameron, R. S., & Rakic, P. (1994). Identification of membrane proteins that comprise the plasmalemmal junction between migrating neurons and radial glial cells. *J. Neurosci., 14*, 3139–3155.

Campbell, M. J., & Morrison, J. H. (1989). Monoclonal antibody to neurofilament protein (SMI-32) labels a subpopulation of pyramidal neurons in the human and monkey neocortex. *J. Comp. Neurol., 282*, 191–205.

Carney, R. S., Bystron, I., Lopez-Bendito, G., & Molnar, Z. (2007). Comparative analysis of extra-ventricular mitoses at early stages of cortical development in rat and human. *Brain Struct. Funct., 212*, 37–54.

Catalano, S. M., Robertson, R. T., & Killackey, H. P. (1996). Individual axon morphology and thalamocortical topography in developing rat somatosensory cortex. *J. Comp. Neurol., 367*, 36–53.

Caviness, V. S., Jr., & Rakic, P. (1978). Mechanisms of cortical development: A view from mutations in mice. *Annu. Rev. Neurosci., 1*, 297–326.

Chalupa, L. M., & Wefers, C. J. (2000). A comparative perspective on the formation of retinal connections in the mammalian brain. In M. S. Gazzaniga (Ed.), *Cognitive neuroscience: A handbook for the field* (pp. 33–43). Cambridge, MA: MIT Press.

Chenn, A., & Walsh, C. A. (2002). Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science, 297*, 365–369.

Chenn, A., & Walsh, C. A. (2003). Increased neuronal production, enlarged forebrains and cytoarchitectural distortions in beta-catenin overexpressing transgenic mice. *Cereb. Cortex, 13*, 599–606.

Chun, J. J., & Shatz, C. J. (1989). Interstitial cells of the adult neocortical white matter are the remnant of the early generated subplate neuron population. *J. Comp. Neurol., 282*, 555–569.

Cobos, I., Puelles, L., & Martinez, S. (2001). The avian telencephalic subpallium originates inhibitory neurons that invade tangentially the pallium (dorsal ventricular ridge and cortical areas). *Dev. Biol., 239*, 30–45.

Cohen-Tannoudji, M., Babinet, C., & Wassef, M. (1994). Early determination of a mouse somatosensory cortex marker. *Nature, 368*, 460–463.

Conel, J. L. (1939). *The postnatal development of the human cerebral cortex: I. The cortex of the newborn*. Cambridge, MA: Harvard University Press.

Creutzfeldt, O. D. (1977). Generality of the functional structure of the neocortex. *Naturwissenschaften, 64*, 507–517.

Crowley, J. C., & Katz, L. C. (2000). Early development of ocular dominance columns. *Science, 290*, 1321–1324.

deAzevedo, L. C., Fallet, C., Moura-Neto, V., Daumas-Duport, C., Hedin-Pereira, C., & Lent, R. (2003). Cortical radial glial cells in human fetuses: Depth-correlated transformation into astrocytes. *J. Neurobiol., 55*, 288–298.

De Carlos, J. A., Lopez-Mascaraque, L., & Valverde, F. (1996). Dynamics of cell migration from the lateral ganglionic eminence in the rat. *J. Neurosci., 16*, 6146–6156.

De Carlos, J. A., & O'Leary, D. D. (1992). Growth and targeting of subplate axons and establishment of major cortical pathways. *J. Neurosci., 12*, 1194–1211.

DeFelipe, J., Alonso-Nanclares, L., & Arellano, J. I. (2002). Microstructure of the neocortex: Comparative aspects. *J. Neurocytol., 31*, 299–316.

DeFelipe, J., Hendry, S. H., Hashikawa, T., Molinari, M., & Jones, E. G. (1990). A microcolumnar structure of monkey cerebral cortex revealed by immunocytochemical studies of double bouquet cell axons. *Neuroscience, 37*, 655–673.

Dehay, C., Giroud, P., Berland, M., Smart, I., & Kennedy, H. (1993). Modulation of the cell cycle contributes to the parcellation of the primate visual cortex. *Nature, 366*, 464–466.

Del Rio, M. R., & DeFelipe, J. (1996). Colocalization of calbindin D-28k, calretinin, and GABA immunoreactivities in neurons of the human temporal cortex. *J. Comp. Neurol., 369*, 472–482.

des Portes, V., Pinard, J. M., Billuart, P., Vinet, M. C., Koulakoff, A., Carrie, A., et al. (1998). A novel CNS gene required for neuronal migration and involved in X-linked subcortical laminar heterotopia and lissencephaly syndrome. *Cell, 92*, 51–61.

Donoghue, M. J., & Rakic, P. (1999). Molecular evidence for the early specification of presumptive functional domains in the embryonic primate cerebral cortex. *J. Neurosci., 19*, 5967–5979.

Easter, S. S., Jr., Purves, D., Rakic, P., & Spitzer, N. C. (1985). The changing view of neural specificity. *Science, 230*, 507–511.

Eccles, J. C. (1984). The cerebral neocortex: A theory of its operation. In E. G. Jones & A. Peters (Eds.), *Cerebral cortex* (pp. 1–36). New York: Plenum Press.

Ellis, R. E., & Horvitz, H. R. (1991). Two *C. elegans* genes control the programmed deaths of specific cells in the pharynx. *Development, 112*, 591–603.

Elston, G. N. (2003). Cortex, cognition and the cell: New insights into the pyramidal neuron and prefrontal function. *Cereb. Cortex, 13*, 1124–1138.

Elston, G. N., Benavides-Piccione, R., & DeFelipe, J. (2001). The pyramidal cell in cognition: A comparative study in human and monkey. *J. Neurosci., 21*, RC163.

Elston, G. N., Benavides-Piccione, R., Elston, A., Zietsch, B., DeFelipe, J., Manger, P., et al. (2006). Specializations of the granular prefrontal cortex of primates: Implications for cognitive processing. *Anat. Rec. A Discov. Mol. Cell Evol. Biol., 288*, 26–35.

Emerling, D. E., & Lander, A. D. (1994). Laminar specific attachment and neurite outgrowth of thalamic neurons on cultured slices of developing cerebral neocortex. *Development, 120*, 2811–2822.

Feng, Y., & Walsh, C. A. (2001). Protein-protein interactions, cytoskeletal regulation and neuronal migration. *Nat. Rev. Neurosci., 2*, 408–416.

Ferri, R. T., & Levitt, P. (1993). Cerebral cortical progenitors are fated to produce region-specific neuronal populations. *Cereb. Cortex, 3*, 187–198.

Fishell, G., & Kriegstein, A. R. (2003). Neurons from radial glia: The consequences of asymmetric inheritance. *Curr. Opin. Neurobiol., 13*, 34–41.

Frantz, G. D., Bohner, A. P., Akers, R. M., & McConnell, S. K. (1994). Regulation of the POU domain gene SCIP during cerebral cortical development. *J. Neurosci.*, *14*, 472–485.

Fukuchi-Shimogori, T., & Grove, E. A. (2001). Neocortex patterning by the secreted signaling molecule FGF8. *Science*, *294*, 1071–1074.

Gal, J. S., Morozov, Y. M., Ayoub, A. E., Chatterjee, M., Rakic, P., & Haydar, T. F. (2006). Molecular and morphological heterogeneity of neural precursors in the mouse neocortical proliferative zones. *J. Neurosci.*, *26*, 1045–1056.

Gazzaniga, M. S. (2008). *Human: The science behind what makes us unique*. New York: HarperCollins.

Ge, W., He, F., Kim, K. J., Blanchi, B., Coskun, V., Nguyen, L., et al. (2006). Coupling of cell migration with neurogenesis by proneural bHLH factors. *Proc. Natl. Acad. Sci. USA*, *103*, 1319–1324.

Gleeson, J. G., Allen, K. M., Fox, J. W., Lamperti, E. D., Berkovic, S., Scheffer, I., et al. (1998). Doublecortin, a brain-specific gene mutated in human X-linked lissencephaly and double cortex syndrome, encodes a putative signaling protein. *Cell*, *92*, 63–72.

Gleeson, J. G., Lin, P. T., Flanagan, L. A., & Walsh, C. A. (1999). Doublecortin is a microtubule-associated protein and is expressed widely by migrating neurons. *Neuron*, *23*, 257–271.

Gleeson, J. G., & Walsh, C. A. (2000). Neuronal migration disorders: From genetic diseases to developmental mechanisms. *Trends Neurosci.*, *23*, 352–359.

Glucksmann, A. (1952). The response of human tissues to radiation with special reference to differentiation. *Br. J. Radiol.*, *25*, 38–43.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum & V. Mountcastle (Eds.), *Handbook of physiology* (pp. 373–417). Bethesda, MD: American Physiology Society.

Goldman-Rakic, P. S., & Rakic, P. (1984). Experimental modification of gyral patterns. In N. Geschwind & A. M. Galaburda (Eds.), *Cerebral dominance: The biological foundation* (pp. 179–192). Cambridge, MA: Harvard University Press.

Gregory, W. A., Edmondson, J. C., Hatten, M. E., & Mason, C. A. (1988). Cytology and neuron-glial apposition of migrating cerebellar granule cells in vitro. *J. Neurosci.*, *8*, 1728–1738.

Guillemot, F. (2007). Spatial and temporal specification of neural fates by transcription factor codes. *Development*, *134*, 3771–3780.

Hartfuss, E., Galli, R., Heins, N., & Gotz, M. (2001). Characterization of CNS precursor subtypes and radial glia. *Dev. Biol.*, *229*, 15–30.

Hatten, M. E. (2002). New directions in neuronal migration. *Science*, *297*, 1660–1663.

Hatten, M. E., & Mason, C. A. (1990). Mechanisms of glial-guided neuronal migration in vitro and in vivo. *Experientia*, *46*, 907–916.

Hattori, M., Adachi, H., Tsujimoto, M., Arai, H., & Inoue, K. (1994). Miller-Dieker lissencephaly gene encodes a subunit of brain platelet–activating factor acetylhydrolase [corrected]. *Nature*, *370*, 216–218.

Haydar, T. F., Kuan, C. Y., Flavell, R. A., & Rakic, P. (1999). The role of cell death in regulating the size and shape of the mammalian forebrain. *Cereb. Cortex*, *9*, 621–626.

He, W., Ingraham, C., Rising, L., Goderie, S., & Temple, S. (2001). Multipotent stem cells from the mouse basal forebrain contribute GABAergic neurons and oligodendrocytes to the cerebral cortex during embryogenesis. *J. Neurosci.*, *21*, 8854–8862.

Hendry, S. H., Schwark, H. D., Jones, E. G., & Yan, J. (1987). Numbers and proportions of GABA-immunoreactive neurons in different areas of monkey cerebral cortex. *J. Neurosci.*, *7*, 1503–1519.

Higginbotham, H. R., & Gleeson, J. G. (2007). The centrosome in neuronal development. *Trends Neurosci.*, *30*, 276–283.

His, W. (1874). *Unserer Körperform und das physiologische Problem ihrer Enstehung*. Leipzig: Engelman.

His, W. (1904). *Die Entwicklung des menschlichen Gehirns während der ersten Monate*. Leipzig: Hirzel.

Jones, E. G. (2000). Microcolumns in the cerebral cortex. *Proc. Natl. Acad. Sci. USA*, *97*, 5019–5021.

Kaiserman-Abramof, I. R., Graybiel, A. M., & Nauta, W. J. (1980). The thalamic projection to cortical area 17 in a congenitally anophthalmic mouse strain. *Neuroscience*, *5*, 41–52.

Kanold, P. O., Kara, P., Reid, R. C., & Shatz, C. J. (2003). Role of subplate neurons in functional maturation of visual cortical columns. *Science*, *301*, 521–525.

Kennedy, H., & Dehay, C. (1988). Functional implications of the anatomical organization of the callosal projections of visual areas V1 and V2 in the macaque monkey. *Behav. Brain Res.*, *29*, 225–236.

Kennedy, H., & Dehay, C. (1993). Cortical specification of mice and men. *Cereb. Cortex*, *3*, 171–186.

Koketsu, D., Mikami, A., Miyamoto, Y., & Hisatsune, T. (2003). Nonrenewal of neurons in the cerebral neocortex of adult macaque monkeys. *J. Neurosci.*, *23*, 937–942.

Komuro, H., & Rakic, P. (1992). Selective role of N-type calcium channels in neuronal migration. *Science*, *257*, 806–809.

Komuro, H., & Rakic, P. (1993). Modulation of neuronal migration by NMDA receptors. *Science*, *260*, 95–97.

Komuro, H., & Rakic, P. (1996). Intracellular Ca2+ fluctuations modulate the rate of neuronal migration. *Neuron*, *17*, 275–285.

Kornack, D. R., & Rakic, P. (1995). Radial and horizontal deployment of clonally related cells in the primate neocortex: Relationship to distinct mitotic lineages. *Neuron*, *15*, 311–321.

Kornack, D. R., & Rakic, P. (1999). Continuation of neurogenesis in the hippocampus of the adult macaque monkey. *Proc. Natl. Acad. Sci. USA*, *96*, 5768–5773.

Kornack, D. R., & Rakic, P. (2001a). Cell proliferation without neurogenesis in adult primate neocortex. *Science*, *294*, 2127–2130.

Kornack, D. R., & Rakic, P. (2001b). The generation, migration, and differentiation of olfactory neurons in the adult primate brain. *Proc. Natl. Acad. Sci. USA*, *98*, 4752–4757.

Kostovic, I., & Goldman-Rakic, P. S. (1983). Transient cholinesterase staining in the mediodorsal nucleus of the thalamus and its connections in the developing human and monkey brain. *J. Comp. Neurol.*, *219*, 431–447.

Kostovic, I., & Molliver, D. E. (1974). A new interpretation of the laminar development of cerebral cortex: Synaptogenesis in different layers of neopalium in the human fetus. *Anat. Rec.*, *178*, 395.

Kostovic, I., & Rakic, P. (1980). Cytology and time of origin of interstitial neurons in the white matter in infant and adult human and monkey telencephalon. *J. Neurocytol.*, *9*, 219–242.

Kostovic, I., & Rakic, P. (1984). Development of prestriate visual projections in the monkey and human fetal cerebrum revealed by transient cholinesterase staining. *J. Neurosci.*, *4*, 25–42.

Kostovic, I., & Rakic, P. (1990). Developmental history of the transient subplate zone in the visual and somatosensory cortex of the macaque monkey and human brain. *J. Comp. Neurol.*, *297*, 441–470.

Kuan, C. Y., Elliott, E. A., Flavell, R. A., & Rakic, P. (1997). Restrictive clonal allocation in the chimeric mouse brain. *Proc. Natl. Acad. Sci. USA*, *94*, 3374–3379.

Kuida, K., Haydar, T. F., Kuan, C. Y., Gu, Y., Taya, C., Karasuyama, H., et al. (1998). Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell*, *94*, 325–337.

Kuida, K., Zheng, T. S., Na, S., Kuan, C., Yang, D., Karasuyama, H., et al. (1996). Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature*, *384*, 368–372.

Kuljis, R. O., & Rakic, P. (1990). Hypercolumns in primate visual cortex can develop in the absence of cues from photoreceptors. *Proc. Natl. Acad. Sci. USA*, *87*, 5303–5306.

LaMantia, A. S., & Rakic, P. (1990). Axon overproduction and elimination in the corpus callosum of the developing rhesus monkey. *J. Neurosci.*, *10*, 2156–2175.

Lavdas, A. A., Grigoriou, M., Pachnis, V., & Parnavelas, J. G. (1999). The medial ganglionic eminence gives rise to a population of early neurons in the developing cerebral cortex. *J. Neurosci.*, *19*, 7881–7888.

Letinic, K., Zoncu, R., & Rakic, P. (2002). Origin of GABAergic neurons in the human neocortex. *Nature*, *417*, 645–649.

Levitt, P., Barbe, M. F., & Eagleson, K. L. (1997). Patterning and specification of the cerebral cortex. *Annu. Rev. Neurosci.*, *20*, 1–24.

Levitt, P., Cooper, M. L., & Rakic, P. (1981). Coexistence of neuronal and glial precursor cells in the cerebral ventricular zone of the fetal monkey: An ultrastructural immunoperoxidase analysis. *J. Neurosci.*, *1*, 27–39.

Levitt, P., & Rakic, P. (1980). Immunoperoxidase localization of glial fibrillary acidic protein in radial glial cells and astrocytes of the developing rhesus monkey brain. *J. Comp. Neurol.*, *193*, 815–840.

Liao, B. Y., & Zhang, J. (2008). Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc. Natl. Acad. Sci. USA*, *105*, 6987–6992.

Lidow, M. S., Goldman-Rakic, P. S., & Rakic, P. (1991). Synchronized overproduction of neurotransmitter receptors in diverse regions of the primate cerebral cortex. *Proc. Natl. Acad. Sci. USA*, *88*, 10218–10221.

Lidow, M. S., & Rakic, P. (1994). Unique profiles of the alpha 1-, alpha 2-, and beta-adrenergic receptors in the developing cortical plate and transient embryonic zones of the rhesus monkey. *J. Neurosci.*, *14*, 4064–4078.

Lois, C., & Alvarez-Buylla, A. (1994). Long-distance neuronal migration in the adult mammalian brain. *Science*, *264*, 1145–1148.

LoTurco, J. J., Owens, D. F., Heath, M. J., Davis, M. B., & Kriegstein, A. R. (1995). GABA and glutamate depolarize cortical progenitor cells and inhibit DNA synthesis. *Neuron*, *15*, 1287–1298.

Luskin, M. B., Pearlman, A. L., & Sanes, J. R. (1988). Cell lineage in the cerebral cortex of the mouse studied in vivo and in vitro with a recombinant retrovirus. *Neuron*, *1*, 635–647.

Luskin, M. B., & Shatz, C. J. (1985). Studies of the earliest generated cells of the cat's visual cortex: Cogeneration of subplate and marginal zones. *J. Neurosci.*, *5*, 1062–1075.

Malatesta, P., Hack, M. A., Hartfuss, E., Kettenmann, H., Klinkert, W., Kirchhoff, F., et al. (2003). Neuronal or glial progeny: Regional differences in radial glia fate. *Neuron*, *37*, 751–764.

Malatesta, P., Hartfuss, E., & Gotz, M. (2000). Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage. *Development*, *127*, 5253–5263.

Marin, O., & Rubenstein, J. L. (2001). A long, remarkable journey: Tangential migration in the telencephalon. *Nat. Rev. Neurosci.*, *2*, 780–790.

Marino, L., Connor, R. C., Fordyce, R. E., Herman, L. M., Hof, P. R., Lefebvre, L., et al. (2007). Cetaceans have complex brains for complex cognition. *PLoS Biol.*, *5*, e139.

Marin-Padilla, M. (1988). Early ontogenesis of the human cerebral cortex. In E. G. Jones & A. Peters (Eds.), *Cerebral cortex development and maturation of cerebral cortex* (pp. 1–34). New York: Plenum Press.

McConnell, S. K. (1988). Development and decision-making in the mammalian cerebral cortex. *Brain Res.*, *472*, 1–23.

McConnell, S. K., Ghosh, A., & Shatz, C. J. (1994). Subplate pioneers and the formation of descending connections from cerebral cortex. *J. Neurosci.*, *14*, 1892–1907.

McConnell, S. K., & Kaznowski, C. E. (1991). Cell cycle dependence of laminar determination in developing neocortex. *Science*, *254*, 282–285.

Meinecke, D. L., & Peters, A. (1987). GABA immunoreactive neurons in rat visual cortex. *J. Comp. Neurol.*, *261*, 388–404.

Meissirel, C., Wikler, K. C., Chalupa, L. M., & Rakic, P. (1997). Early divergence of magnocellular and parvocellular functional subsystems in the embryonic primate visual system. *Proc. Natl. Acad. Sci. USA*, *94*, 5900–5905.

Menezes, J. R., & Luskin, M. B. (1994). Expression of neuron-specific tubulin defines a novel population in the proliferative layers of the developing telencephalon. *J. Neurosci.*, *14*, 5399–5416.

Merkle, F. T., Mirzadeh, Z., & Alvarez-Buylla, A. (2007). Mosaic organization of neural stem cells in the adult brain. *Science*, *317*, 381–384.

Micheva, K. D., & Beaulieu, C. (1995). An anatomical substrate for experience-dependent plasticity of the rat barrel field cortex. *Proc. Natl. Acad. Sci. USA*, *92*, 11834–11838.

Miyashita-Lin, E. M., Hevner, R., Wassarman, K. M., Martinez, S., & Rubenstein, J. L. (1999). Early neocortical regionalization in the absence of thalamic innervation. *Science*, *285*, 906–909.

Mizutani, K., Yoon, K., Dang, L., Tokunaga, A., & Gaiano, N. (2007). Differential Notch signalling distinguishes neural stem cells from intermediate progenitors. *Nature*, *449*, 351–355.

Molliver, M. E., Kostovic, I., & van der Loos, H. (1973). The development of synapses in cerebral cortex of the human fetus. *Brain Res.*, *50*, 403–407.

Molyneaux, B. J., Arlotta, P., Menezes, J. R., & Macklis, J. D. (2007). Neuronal subtype specification in the cerebral cortex. *Nat. Rev. Neurosci.*, *8*, 427–437.

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain*, *120* (Pt. 4), 701–722.

Nadarajah, B., Alifragis, P., Wong, R. O., & Parnavelas, J. G. (2003). Neuronal migration in the developing cerebral cortex: Observations based on real-time imaging. *Cereb. Cortex*, *13*, 607–611.

Nakatsuji, M., Kadokawa, Y., & Suemori, H. (1991). Radial columnar patches in the chimeric cerebral cortex visualized by

use of mouse embryonic stem cells expressing Beta-galactosidase. *Dev. Growth Differ.*, *33*, 571–578.

Nimchinsky, E. A., Gilissen, E., Allman, J. M., Perl, D. P., Erwin, J. M., & Hof, P. R. (1999). A neuronal morphologic type unique to humans and great apes. *Proc. Natl. Acad. Sci. USA*, *96*, 5268–5273.

Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S., & Kriegstein, A. R. (2001). Neurons derived from radial glial cells establish radial units in neocortex. *Nature*, *409*, 714–720.

Noctor, S. C., Flint, A. C., Weissman, T. A., Wong, W. S., Clinton, B. K., & Kriegstein, A. R. (2002). Dividing precursor cells of the embryonic cortical ventricular zone have morphological and molecular characteristics of radial glia. *J. Neurosci.*, *22*, 3161–3173.

Olavarria, J., & van Sluyters, R. C. (1984). Callosal connections of the posterior neocortex in normal-eyed, congenitally anophthalmic, and neonatally enucleated mice. *J. Comp. Neurol.*, *230*, 249–268.

Oppenheim, R. W. (1991). Cell death during development of the nervous system. *Annu. Rev. Neurosci.*, *14*, 453–501.

Parnavelas, J. G., Barfield, J. A., Franke, E., & Luskin, M. B. (1991). Separate progenitor cells give rise to pyramidal and nonpyramidal neurons in the rat telencephalon. *Cereb. Cortex*, *1*, 463–468.

Petanjek, Z., Berger, B., & Esclapez, M. (2009). Origins of cortical GABAergic neurons in the cynomolgus monkey. *Cereb. Cortex*, *19*, 249–262.

Petanjek, Z., Dujmovic, A., Kostovic, I., & Esclapez, M. (2008). Distinct origin of GABA-ergic neurons in forebrain of man, nonhuman primates and lower mammals. *Collegium Antropologicum*, *32*, Suppl. 1, 9–17.

Poliakov, G. I. (1959). Progressive neuron differentiation of the human cerebral cortex in ontogenesis. In S. A. Sarkisov & S. N. Preobrazenskaya (Eds.), *Development of the central nervous system* (in Russian) (pp. 11–26). Moscow: Medgiz.

Poliakov, G. I. (1965). Development of the cerebral neocortex during the first half of intrauterine life. In S. A. Sarkisov (Ed.), *Development of the child's brain* (in Russian) (pp. 22–52). Leningrad: Medicinam.

Poluch, S., & Juliano, S. L. (2007). A normal radial glial scaffold is necessary for migration of interneurons during neocortical development. *Glia*, *55*, 822–830.

Porteus, M. H., Bulfone, A., Ciaranello, R. D., & Rubenstein, J. L. (1991). Isolation and characterization of a novel cDNA clone encoding a homeodomain that is developmentally regulated in the ventral forebrain. *Neuron*, *7*, 221–229.

Rakic, P. (1971). Neuron-glia relationship during granule cell migration in developing cerebellar cortex: A Golgi and electronmicroscopic study in *Macacus rhesus*. *J. Comp. Neurol.*, *141*, 283–312.

Rakic, P. (1972). Mode of cell migration to the superficial layers of fetal monkey neocortex. *J. Comp. Neurol.*, *145*, 61–83.

Rakic, P. (1973). Kinetics of proliferation and latency between final cell division and onset of differentiation of cerebellar stellate and basket neurons. *J. Comp. Neurol.*, *147*, 523–546.

Rakic, P. (1974). Neurons in rhesus monkey visual cortex: Systematic relation between time of origin and eventual disposition. *Science*, *183*, 425–427.

Rakic, P. (1975). Timing of major ontogenetic events in the visual cortex of the rhesus monkey. In N. A. Buchwald & M. Brazier (Eds.), *Brain mechanisms in mental retardation* (pp. 3–40). New York: Academic Press.

Rakic, P. (1976a). Differences in the time of origin and in eventual distribution of neurons in areas 17 and 18 of the visual cortex in the rhesus monkey. *Exp. Brain Res.*, Suppl. 1, 244–248.

Rakic, P. (1976b). Prenatal genesis of connections subserving ocular dominance in the rhesus monkey. *Nature*, *261*, 467–471.

Rakic, P. (1977). Prenatal development of the visual system in rhesus monkey. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *278*, 245–260.

Rakic, P. (1978). Neuronal migration and contact guidance in the primate telencephalon. *Postgrad. Med. J.*, *54*, Suppl. 1, 25–40.

Rakic, P. (1981). Neuronal-glial interaction during brain development. *Trends Neurosci.*, *4*, 184–187.

Rakic, P. (1985). Limits of neurogenesis in primates. *Science*, *227*, 1054–1056.

Rakic, P. (1988). Specification of cerebral cortical areas. *Science*, *241*, 170–176.

Rakic, P. (1990). Principles of neural cell migration. *Experientia*, *46*, 882–891.

Rakic, P. (1995a). Radial versus tangential migration of neuronal clones in the developing cerebral cortex. *Proc. Natl. Acad. Sci. USA*, *92*, 11323–11327.

Rakic, P. (1995b). A small step for the cell, a giant leap for mankind: A hypothesis of neocortical expansion during evolution. *Trends Neurosci.*, *18*, 383–388.

Rakic, P. (2000). Advantages of the mouse model: From spontaneous to induced mutations. In A. Goffinet & P. Rakic (Eds.), *The mouse brain development* (pp. 1–19). Berlin: Springer-Verlag.

Rakic, P. (2002). Neurogenesis in adult primate neocortex: An evaluation of the evidence. *Nat. Rev. Neurosci.*, *3*, 65–71.

Rakic, P. (2003). Elusive radial glial cells: Historical and evolutionary perspective. *Glia*, *43*, 19–32.

Rakic, P. (2006). Neuroscience: No more cortical neurons for you. *Science*, *313*, 928–929.

Rakic, P. (2007). The radial edifice of cortical architecture: From neuronal silhouettes to genetic engineering. *Brain Res. Brain Res. Rev.*, *55*, 204–219.

Rakic, P., Ayoub, A. E., Dominguez, M., & Breunig, J. J. (in press). Decision by division: Making cortical maps. *Trends Neurosci.*

Rakic, P., Cameron, R. S., & Komuro, H. (1994). Recognition, adhesion, transmembrane signaling and cell motility in guided neuronal migration. *Curr. Opin. Neurobiol.*, *4*, 63–69.

Rakic, P., Knyihar-Csillik, E., & Csillik, B. (1996). Polarity of microtubule assemblies during neuronal cell migration. *Proc. Natl. Acad. Sci. USA*, *93*, 9218–9222.

Rakic, P., & Komuro, H. (1995). The role of receptor/channel activity in neuronal cell migration. *J. Neurobiol.*, *26*, 299–315.

Rakic, P., & Lidow, M. S. (1995). Distribution and density of monoamine receptors in the primate visual cortex devoid of retinal input from early embryonic stages. *J. Neurosci.*, *15*, 2561–2574.

Rakic, P., & Nowakowski, R. S. (1981). The time of origin of neurons in the hippocampal region of the rhesus monkey. *J. Comp. Neurol.*, *196*, 99–128.

Rakic, P., & Riley, K. P. (1983a). Overproduction and elimination of retinal axons in the fetal rhesus monkey. *Science*, *219*, 1441–1444.

Rakic, P., & Riley, K. P. (1983b). Regulation of axon number in primate optic nerve by prenatal binocular competition. *Nature*, *305*, 135–137.

Rakic, P., & Sidman, R. L. (1968). Supravital DNA synthesis in the developing human and mouse brain. *J. Neuropathol. Exp. Neurol.*, *27*, 246–276.

Rakic, P., & Singer, W. (1988). *Neurobiology of the neocortex*. New York: John Wiley & Sons.

Rakic, P., Suner, I., & Williams, R. W. (1991). A novel cytoarchitectonic area induced experimentally within the primate visual cortex. *Proc. Natl. Acad. Sci. USA*, *88*, 2083–2087.

Rakic, S., & Zecevic, N. (2003). Emerging complexity of layer I in human cerebral cortex. *Cereb. Cortex*, *13*, 1072–1083.

Rasin, M. R., Gazula, V. R., Breunig, J. J., Kwan, K. Y., Johnson, M. B., Liu-Chen, S., et al. (2007). Numb and Numbl are required for maintenance of cadherin-based adhesion and polarity of neural progenitors. *Nat. Neurosci.*, *10*, 819–827.

Reid, C. B., Liang, I., & Walsh, C. (1995). Systematic widespread clonal organization in cerebral cortex. *Neuron*, *15*, 299–310.

Richards, L. J., Koester, S. E., Tuttle, R., & O'Leary, D. D. (1997). Directed growth of early cortical axons is influenced by a chemoattractant released from an intermediate target. *J. Neurosci.*, *17*, 2445–2458.

Rivas, R. J., & Hatten, M. E. (1995). Motility and cytoskeletal organization of migrating cerebellar granule neurons. *J. Neurosci.*, *15*, 981–989.

Sanes, J. R. (1989). Analysing cell lineage with a recombinant retrovirus. *Trends Neurosci.*, *12*, 21–28.

Schaar, B. T., & McConnell, S. K. (2005). Cytoskeletal coordination during neuronal migration. *Proc. Natl. Acad. Sci. USA*, *102*, 13652–13657.

Schachner, M., Faissner, A., Fischer, G., Keilhauer, G., Kruse, J., Ktinemund, V., Lindner, J., & Wernecke, H. (1985). Functional and structural aspects of the cell surface in mammalian nervous system development. In G. M. Edelman, W. E. Gall, & J. P. Thiery (Eds.), *The cell in contact: Adhesions and junctions as morphogenetic determinants*. (pp. 257–275). New York: John Wiley & Sons.

Schmechel, D. E., & Rakic, P. (1979). Arrested proliferation of radial glial cells during midgestation in rhesus monkey. *Nature*, *277*, 303–305.

Schwartz, M. L., Rakic, P., & Goldman-Rakic, P. S. (1991). Early phenotype expression of cortical neurons: Evidence that a subclass of migrating neurons have callosal axons. *Proc. Natl. Acad. Sci. USA*, *88*, 1354–1358.

Sestan, N., Rakic, P., & Donoghue, M. J. (2001). Independent parcellation of the embryonic visual cortex and thalamus revealed by combinatorial Eph/ephrin gene expression. *Curr. Biol.*, *11*, 39–43.

Shimojo, H., Ohtsuka, T., & Kageyama, R. (2008). Oscillations in notch signaling regulate maintenance of neural progenitors. *Neuron*, *58*, 52–64.

Sidman, R. L., & Rakic, P. (1973). Neuronal migration, with special reference to developing human brain: A review. *Brain Res.*, *62*, 1–35.

Sidman, R. L., & Rakic, P. (1982). Development of the human central nervous system. In W. Haymaker & R. D. Adams (Eds.), *Histology and histopathology of the nervous system* (pp. 3–145). Springfield, IL: Charles C. Thomas.

Smart, I. H., Dehay, C., Giroud, P., Berland, M., & Kennedy, H. (2002). Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. *Cereb. Cortex*, *12*, 37–53.

Solecki, D. J., Model, L., Gaetz, J., Kapoor, T. M., & Hatten, M. E. (2004). Par6alpha signaling controls glial-guided neuronal migration. *Nat. Neurosci.*, *7*, 1195–1203.

Soriano, E., Dumesnil, N., Auladell, C., Cohen-Tannoudji, M., & Sotelo, C. (1995). Molecular heterogeneity of progenitors and radial migration in the developing cerebral cortex revealed by transgene expression. *Proc. Natl. Acad. Sci. USA*, *92*, 11676–11680.

Szentagothai, J. (1978). The Ferrier Lecture, 1977. The neuron network of the cerebral cortex: A functional interpretation. *Proc. R. Soc. Lond. B Biol. Sci.*, *201*, 219–248.

Tamamaki, N., Fujimori, K. E., & Takauji, R. (1997). Origin and route of tangentially migrating neurons in the developing neocortical intermediate zone. *J. Neurosci.*, *17*, 8313–8323.

Tamamaki, N., Nakamura, K., Okamoto, K., & Kaneko, T. (2001). Radial glia is a progenitor of neocortical neurons in the developing cerebral cortex. *Neurosci. Res.*, *41*, 51–60.

Tan, S. S., Kalloniatis, M., Sturm, K., Tam, P. P., Reese, B. E., & Faulkner-Jones, B. (1998). Separate progenitors for radial and tangential cell dispersion during development of the cerebral neocortex. *Neuron*, *21*, 295–304.

Tramontin, A. D., Garcia-Verdugo, J. M., Lim, D. A., & Alvarez-Buylla, A. (2003). Postnatal development of radial glia and the ventricular zone (VZ): A continuum of the neural stem cell compartment. *Cereb. Cortex*, *13*, 580–587.

Tsai, J. W., Bremner, K. H., & Vallee, R. B. (2007). Dual subcellular roles for LIS1 and dynein in radial neuronal migration in live brain tissue. *Nat. Neurosci.*, *10*, 970–979.

Tsai, L. H., & Gleeson, J. G. (2005). Nucleokinesis in neuronal migration. *Neuron*, *46*, 383–388.

Verhage, M., Maia, A. S., Plomp, J. J., Brussaard, A. B., Heeroma, J. H., Vermeer, H., et al. (2000). Synaptic assembly of the brain in the absence of neurotransmitter secretion. *Science*, *287*, 864–869.

Williams, B. P., & Price, J. (1995). Evidence for multiple precursor cell types in the embryonic rat cerebral cortex. *Neuron*, *14*, 1181–1188.

Williams, R. W., & Herrup, K. (1988). The control of neuron number. *Annu. Rev. Neurosci.*, *11*, 423–453.

Yanez, I. B., Munoz, A., Contreras, J., Gonzalez, J., Rodriguez-Veiga, E., & DeFelipe, J. (2005). Double bouquet cell in the human cerebral cortex and a comparison with other mammals. *J. Comp. Neurol.*, *486*, 344–360.

Yoon, K. J., Koo, B. K., Im, S. K., Jeong, H. W., Ghim, J., Kwon, M. C., et al. (2008). Mind bomb 1–expressing intermediate progenitors generate notch signaling to maintain radial glial cells. *Neuron*, *58*, 519–531.

Zecevic, N. (1998). Synaptogenesis in layer I of the human cerebral cortex in the first half of gestation. *Cereb. Cortex*, *8*, 245–252.

Zecevic, N., Chen, Y., & Filipovic, R. (2005). Contributions of cortical subventricular zone to the development of the human cerebral cortex. *J. Comp. Neurol.*, *491*, 109–122.

Zhong, W., Feder, J. N., Jiang, M. M., Jan, L. Y., & Jan, Y. N. (1996). Asymmetric localization of a mammalian numb homolog during mouse cortical neurogenesis. *Neuron*, *17*, 43–53.

2 Early Development of Neuronal Circuitry of the Human Prefrontal Cortex

IVICA KOSTOVIĆ AND MILOŠ JUDAŠ

ABSTRACT The early development of cortical circuitry provides the biological substrate for human cognitive and psychological maturation. Neuronal circuitry of the human frontal cortex appears around the eighth postconceptional week (8 PCW) with two synaptic strata, engagement of fetal neurons, and spontaneous activity. Midfetal and late-fetal cortex show transient lamination, differentiation of subplate, deep synaptogenesis, and growth of thalamocortical afferents. Cortical interaction with thalamic afferents occurs around 24 PCW. Late preterm and neonatal periods are characterized by growth of cortico-cortical fibers and resolution of transient circuitry. During early infancy (2–6 months) rapid synaptogenesis coincides with reorganization of cortico-cortical pathways. Initial environmentally driven "cognitive" circuitry consists of increased number of synapses, differentiated layer V pyramids, "dormant" layer III pyramids and appearance of inhibitory neurons. Full maturity of layer III pyramids, local circuitry neurons and maximal number of synapses is not achieved until 12–24 months when circuitry is layer III "centered" and socially driven. In summary, endogeneous and sensory-driven circuitry develops during prenatal life, initial "cognitive" circuitry appears in late infancy, and maximal number of synapses and full maturity of layer III develop during early childhood.

In revealing the neurobiological basis of cognitive and psychological development in humans, neuroscientists mostly depend on close correspondence between structural, functional, and behavioral features during specific phases of development (Casey, Giedd, & Thomas, 2000; Casey, Tottenham, Liston, & Durston, 2005; Hammock & Levitt, 2006; Kagan & Baird, 2004; Levitt, 2003). Because the neural circuitry that underlies cognitive development is expected to be simpler in the developing than in the adult brain, the developmental approach should provide an easier analysis of principal elements of the neuronal circuitry: synapses, presynaptic axons, and postsynaptic neurons.

It is difficult to correlate structure and function in the developing neuronal circuitry of the frontal lobe because of its complexity (Goldman-Rakic, 1987), life-long maturation (Chugani, Phelps, & Mazziotta, 1987; Fuster, 2002; Giedd et al., 1999; Goldman-Rakic, 1987; Huttenlocher & Dabholkar, 1997; Kostović, Petanjek, Delalle, & Judaš, 1992; Petanjek, Judaš, Kostović, & Uylings, 2008; Sowell et al., 2004) and the existence of human-specific functions (Preuss, 2004; Werker & Vouloumanos, 2001).

Which component of the prefrontal cortical circuitry develops already in utero, that is, before birth? And, knowing that typical prefrontal executive functions develop postnatally, what would be its functional roles during fetal life? In other words, what functions may the prefrontal cortex subserve before the onset of cognition, during the so-called precognition period of life? One would also like to pinpoint qualitative and quantitative features of the maturational status of prefrontal circuitry at the end of the first postnatal year: what level of maturation of prefrontal cortex is required for the onset of cognition and language?

In this review, we summarize the data available on early development of human frontal lobe circuitry, describe new data on neurogenetic events obtained with neuroimaging and fine neurohistological studies, and propose structural criteria for delineation of phases in prefrontal circuitry development. We cover the period from the appearance of the cortical plate and establishment of first cortical synaptic contacts (at 8 PCW) to the third year of life when pyramidal neurons of layer III attain size and complexity greater than those of layer V pyramidal neurons, and language and cognition are already well established (see table 2.1). The following frontal cortical regions will be described: dorsolateral, dorsomedial, orbitomedial, precingulate (area 32), and anterior cingulate (area 24).

Development of neuronal circuitry of the human prefrontal cortex during the early fetal period

STRUCTURAL ORGANIZATION The first cortical cells appear in the neocortical preplate or primordial plexiform layer (Bystron, Rakic, Molnar, & Blakemore, 2006; Bystron,

IVICA KOSTOVIĆ and MILOŠ JUDAŠ Croatian Institute for Brain Research, School of Medicine, University of Zagreb, Zagreb, Croatia

TABLE 2.1

Early development of human frontal lobe circuitry

Period	Age Range	Description of Circuitry	Type of Circuitry	Type of Activity
Embryonic	6–7 PCW	Nonsynaptic preplate network	Oscillating, spontaneous	Endogenous
Early fetal	8–14 PCW	Two synaptic strata, in SP and MZ; monoaminergic afferents; regional differences	Transient spontaneous modulated by monoamines	Endogenous + brain stem and cholinergic basal forebrain
Midfetal and late fetal	15–23 PCW	Transient lamination, prominent SP, and deep synaptogenesis	Transient spontaneous, modulated by extrinsic thalamic afferents	Endogenous
Early preterm	24–32 PCW	The peak of SP, thalamic afferents in CP	Transient circuitry and permanent thalamocortical afferents coexist	Sensory-sensitive asynchrony (EEG), experience independent
Late preterm	33–35 PCW	Synaptogenesis in CP, pyramidal differentiation, cytoarchitectonic belts	Coexistence of increasingly permanent and transient circuitry	Sensory-sensitive synchrony (EEG), experience expectant
Neonatal	1–2 months	Long afferents within the target; layer V pyramids differentiation; ubiquitous layer IV	Permanent circuitry with transient elements	Sensory-driven, layer V centered
Early infancy	2–6 months	Reorganization of corticocortical pathways; rapid synaptogenesis and spinogenesis	Permanent circuitry with resolving transient elements	Sensory-driven; columnar processing
Late infancy	7–12 months	Long connectivity established; layer III pyramids "dormant", areal differentiation; granular-dysgranular	Initial "cognitive" circuitry	Environmentally driven; extrinsic-intrinsic through local circuitry
Early childhood	12–24 months	Maturity of layer III pyramids and local circuits; maximal synaptogenesis	Cognitive	Socially driven; layer III centered

PCW, postconceptional weeks; SP, subplate; MZ, marginal zone; CP, cortical plate.

Blakemore, & Rakic, 2008; Zecevic & Milosevic, 1997) (figure 2.1*A*). Soon thereafter, the formation of the cortical plate (CP) at 8 PCW (figure 2.1*B*) marks the transition from embryonic to fetal period and prefrontal neocortical anlage is now composed of three architectonic zones: marginal zone (MZ), CP, and the presubplate, PSP (Bystron et al., 2008; Kostović & Rakic, 1990). These zones contain two main classes of neurons, radially oriented postmigratory neurons within the CP and randomly oriented and early maturing neurons located above and below the CP, that is, in the MZ and PSP, respectively (Bystron et al., 2008; Kostović, 1990; Marin-Padilla, 1983; Meyer, Schaaps, Moreau, & Goffinet, 2000; Verney, Lebrand, & Gaspar, 2002; Zecevic & Milosevic, 1997).

Between 13 and 14 PCW (the end of the early fetal period, figure 2.1*C*), the deep part of the hitherto densely packed CP transforms into a wider lamina that appears pale in Nissl-stained preparations and merges with PSP, thus forming a new, prominent, and transient zone—the subplate zone (SP) (Kostović & Rakic, 1990; Kostović & Judaš, 2007). This marks the onset of the typical midfetal lamination pattern.

Importantly, the early fetal cerebral wall (telencephalic pallium) shows early regional differences in the thickness and appearance of MZ, CP, and SP across mediolateral, rostrocaudal, and dorsoventral extent. Dorsal fetal pallium, as a forerunner of neocortex, displays thicker and condensed CP, narrow MZ, and wider PSP. The CP becomes thinner toward the interhemispheric fissure where dorsal pallium continues into the medial pallium. The pallium of the cortical limbus (hem) is thin and convoluted with wide MZ, thin and convoluted CP, and almost invisible PSP. The CP is initially not developed in the hippocampal anlage. This early regional specification probably reflects the activity of intrinsic patterning mechanisms (Rakic, 1988, 2006) whereby patterning centers generate, across the dorsal telencephalon, graded expression of transcription factors acting on cortical progenitor cells (Grove & Fukuchi-Shimogori, 2003; O'Leary, Chou, & Sahara, 2007). Thus the frontal telencephalon becomes specified by specific transcription factors, such as the basic fibroblast growth factor (bFGF), expressed in the ventricular zone of the rostral (anterior) telencephalon.

Although this early patterning of the cortical protomap (Rakic, 1988) occurs in humans probably during the second

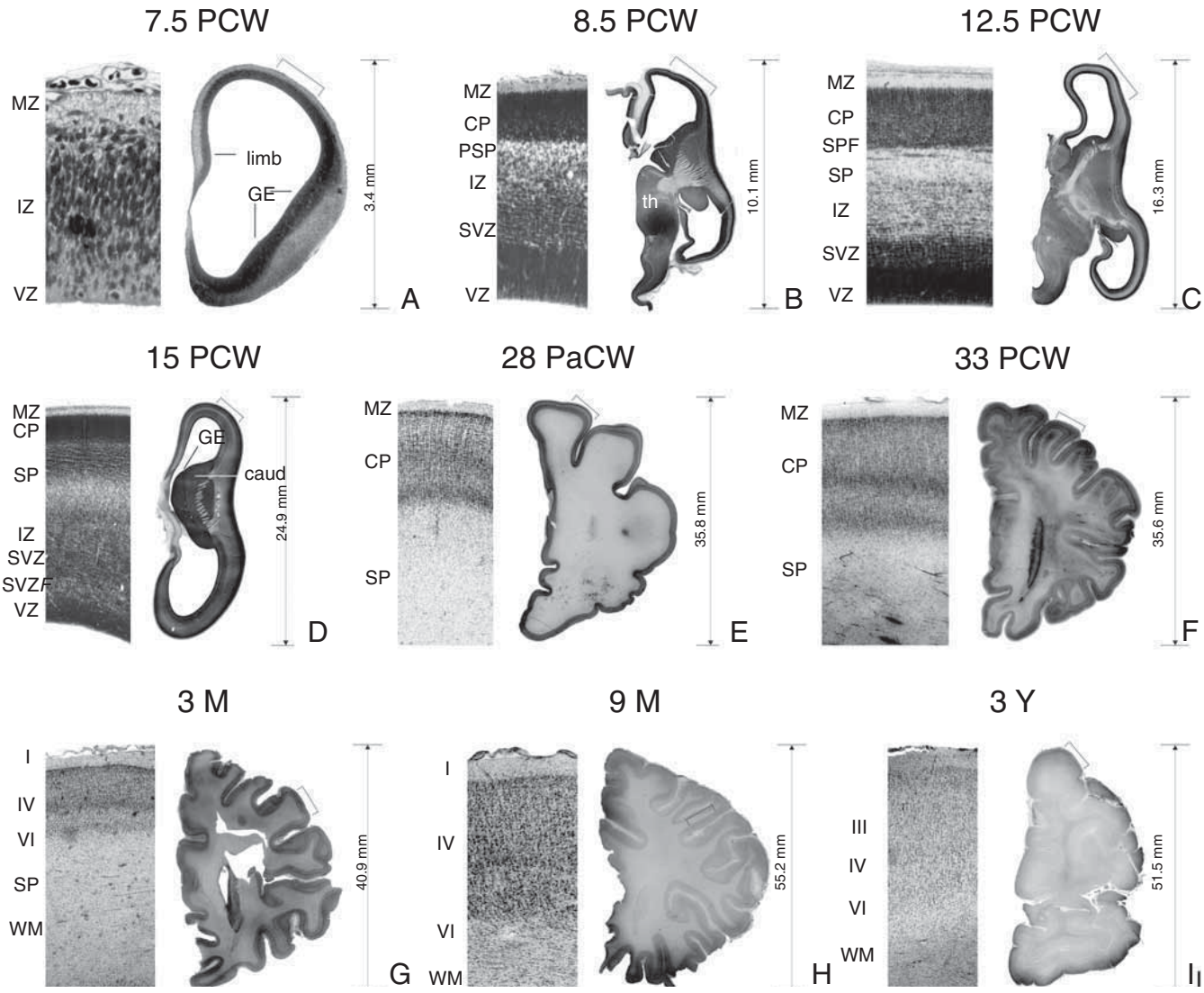


FIGURE 2.1 Development of cytoarchitectonic layers in the prefrontal cortex from the embryonic phase (before the appearance of the cortical plate) to the third year, that is, at 7.5 postconceptional weeks (PCW) (*A*), 8.5 PCW (*B*), 12.5 PCW (*C*), 15 PCW (*D*), 28 PCW (*E*), 33 PCW (*F*), 3 months (*G*), 9 months (*H*), and 3 years (*I*). All layers in prenatal phases are transient, and their laminar changes reflect neurogenetic events: proliferation, migration, differentiation, ingrowth of afferent pathways, and areal differentiation.

Abbreviations for this and subsequent figures: caud, caudate nucleus; CC, corpus callosum; CP, cortical plate; GE, ganglionic eminence; IZ, intermediate zone; limb, limbic (hippocampal pallium); MZ, marginal zone; put, putamen; PSP, presubplate; SP, subplate zone; SPF, the subplate in formation (the "second" cortical plate); SVZ, subventricular zone; SVZ*f*, subventricular fibrillar zone; th, thalamus; VZ, ventricular zone; WM, white matter. Roman numerals (I–VI) correspond to permanent cortical layers; double arrows point to the external capsule.

month of embryonic life—that is, even before the formation of the cortical plate—the specification of cortical areas continues during the fetal period, and thalamic input seems to have a significant role in the final differentiation of cortical areas.

NEUROGENETIC EVENTS Major neurogenetic events are production of young postmitotic neurons in the ventricular zone and their migration through the intermediate zone (figure 2.6). Comparison with data in monkey (Rakic, 1988, 2006) shows that in the early fetal period, neurons destined for superficial (associative) layer III are not born yet. Molecular specification of early cortical neurons was proven using different markers for GABA (Bystron et al., 2006, 2008; Zecevic & Milosevic, 1997; Rakic & Zecevic, 2003) and reelin (Meyer & Goffinet, 1998; Rakic & Zecevic, 2003). For studying early cortical development in a clinical setting it is very important that early transient proliferative, migratory, and synaptic zones were successfully visualized by in vivo, in utero imaging around 13 PCW (Judaš et al., 2005; Kostović, Judaš, Škrablin-Kučić, Štern-Padovan, & Radoš, 2006).

Neuronal Circuitry and Functional Organization The neuronal circuitry of the early prefrontal cortex consists of a small number of synapses, dendrites of postmigratory neurons, and presynaptic axons arising from modulatory extrathalamic subcortical systems. As described for the midlateral telencephalic wall (Molliver, Kostović, & van der Loos, 1973) and the anterior cingulate cortex (Kostović & Krmpotić, 1976) of human fetuses, early synapses display bilaminar distribution and are located above and below the cortical plate. Such bilaminar distribution of early synapses was also demonstrated in equivalent stages (embryonic day 60) of fetal rhesus monkeys (Bourgeois, Goldman-Rakic, & Rakic, 1994).

In the developing human cortex, prospective postsynaptic elements for the early synapses were revealed by Golgi impregnations (Mrzljak, Uylings, Kostović, & van Eden, 1988; Mrzljak, Uylings, van Eden, & Judaš, 1990; Mrzljak, Uylings, Kostović, & van Eden, 1992) and electron microscopy (Kostović & Rakic, 1990), the most likely candidates being neurons of the subplate zone (Kostović & Molliver, 1974; Kostović & Rakic, 1990). However, radially oriented neurons of the immature cortical plate also extend their dendrites in the marginal zone (Marin-Padilla, 1983; Molliver et al., 1973; Mrzljak et al., 1988) and the presubplate (Kostović & Rakic, 1990; Kostović & Judaš, 2007; Mrzljak et al., 1988) and thus represent putative postsynaptic sites for at least some synapses.

There is substantial evidence for the presence of presynaptic axons within early synaptic strata of the human cerebral cortex. Modulatory afferents arrive early from the monoaminergic brainstem tegmentum (Nobin & Björklund, 1973; Zecevic & Verney, 1995) and cholinergic basal forebrain (Kostović, 1986). Although thalamocortical axons are still in the stage of pathway selection, their involvement in the early circuitry cannot be excluded (Allendoerfer & Shatz, 1994; Kostović & Judaš, 2002, 2007; Molliver et al., 1973). Short presynaptic input is also provided by subplate neurons that express glutamate and GABA (Antonini & Shatz, 1990), as well as several neuropeptides (Allendoerfer & Shatz, 1994; Delalle, Evers, Kostović, & Uylings, 1997; Kostović, Štfulj-Fučić, Mrzljak, Jukić, & Delalle, 1991).

Functional Organization At this early age, supragranular neurons involved in long cortico-cortical networking necessary for cognitive functions are not born yet. There are no imaging or functional recording studies of the early human prefrontal cortex. Therefore, we have to rely on experimental studies in animals with similar patterns of circuitry organization. These studies indicate that subplate cells communicate with both nonsynaptic (Albrieux, Platel, Dupuis, Villaz, & Moody, 2004; Dupont, Hanganu, Kilb, Hirsch, & Luhmann, 2006; Voigt, Opitz, & DeLima, 2001) and synaptic contacts (Friauf & Shatz, 1991) and produce spontaneous endogeneous oscillations that do not depend on sensory input. These results suggest that, long before the onset of cognition, the prefrontal cortex of early human fetuses has neuronal circuitry involved in spontaneous endogeneous activity.

Development of neuronal circuitry in the human prefrontal cortex during the midfetal and late fetal period

Structural Organization Transient fetal lamination of the prefrontal cortex (figure 2.1*D*) is characterized by prominent SP (Kostović & Rakic, 1990; Kostović & Judaš, 2007), thick and consolidated CP, and the presence of transient sublayers in the MZ (Kostović, Jovanov-Milošević, Krsnik, Petanjek, & Judaš, 2004/2005). The SP contains an abundant extracellular matrix (Kostović, Judaš, Radoš, & Hrabač, 2002), "waiting" thalamocortical fibers (Kostović & Goldman-Rakic, 1983; Kostović & Rakic, 1984, 1990; Rakic, 1977), and relatively mature polymorphic neurons of quite variable morphology (Kostović, 1990; Mrzljak et al., 1988, 1990, 1992). Transient midfetal zones are easily visualized by both in vitro (Kostović et al., 2002) and in vivo MR imaging (Kostović, Judaš, et al., 2006). The most prominent transient zone is the SP, easily delineated on MRI scans owing to the presence of abundant and hydrophyllic extracellular matrix (Kostović et al., 2002).

During the midfetal and late fetal period the prefrontal cortical areas, as cytoarchitectonically defined in the adult brain by Brodmann (1909) or von Economo and Koskinas (1925), still do not display two defining features, that is, granular layer IV and large pyramidal neurons in sublayer IIIC. At the beginning of this period, layer III neurons are not even born. In addition, the prominent SP determines the lamination pattern, which is very different from that of the postnatal cortex.

However, architectonic differences between major prefrontal cortical regions (dorsolateral, orbitomedial, and orbitolateral) are already visible on Nissl-stained histological sections. Transient fetal zones display regional differences on in vitro MR images of formalin-fixed brains (Kostović, Judaš, et al., 2002) as well as on in vivo, in utero MR images (Kostović, Judaš, et al., 2006). The dorsolateral prefrontal region shows two characteristic features: very thick SP and tangential waves of migratory neurons in the superficial part of the SP, just below the CP.

Neurogenetic Events During the midfetal period, neurons and glial cells are continuously produced in ventricular and subventricular zones (Rakic, 1988, 2006). The subventricular zone is highly developed in the primate brain and seems to be the source of interneurons and glial cells (Bystron et al., 2008; Rakic, 2006). The overall increase in the number of mitotic cycles seems to represent the

cellular basis of prefrontal cortex expansion during human evolution and development (Rakic, 2006).

As neurons are continuously produced, they also continue to migrate through the intermediate zone and SP. Neurons migrating during the midfetal period are destined to supragranular layers and probably correspond to prospective pyramidal neurons of layer III that give rise to associative and commissural cortical pathways (Schwartz, Rakic, & Goldman-Rakic, 1991). The major pattern of migration of principal, pyramidal neurons is radial migration along radial glial cells (Rakic, 2006). This mechanism is particularly important for the thick cerebral wall of the midfetal primate and human brain where neurons have to travel for a long distance from the site of origin in the ventricular-subventricular zones to their final destination in superficial cortical layers. For example, in the midfetal human prefrontal cortex, radially migrating neurons have to traverse 5 to 9 mm thick SP. As seen on Nissl-stained histological sections, massive waves of migratory neurons, together with rows of glial cells, form a transient sublayer within the superficial SP (figure 2.1*D*) of the human prefrontal cortex, which was first noted by von Economo and Koskinas (1925).

During the midfetal period, the growth of subcortico-cortical and cortico-subcortical projection pathways reaches its peak intensity. Thalamocortical pathways grow through the intermediate zone during the pathway selection stage (figure 2.2*A*), and through the SP during the regional target selection stage. The major input for the prefrontal cortex may be traced by selective histochemical staining of growing fiber tracts (Kostović & Goldman-Rakic, 1983) (see also figures 2.3*A*, 2.3*B*). The growth of thalamocortical pathways partially overlaps with continuous growth of afferent fibers from basal forebrain in the external capsule. Fibers from the amygdala were not directly visualized, but this nuclear complex matures very early (Nikolić & Kostović, 1986), and it is very likely that projection to orbitomedial cortex matures in parallel with thalamic projection.

Cortico-subcortical pathways from the prefrontal cortex are more difficult to trace in the human brain. In the rhesus monkey, developing corticostriatal projections are in spatial register with striatal cytoarchitectonic and AChE-rich units (Goldman-Rakic, 1981). In the human striatum, these units develop during midgestation, indicating that midgestation represents a developmental window for growth and establishment of human prefrontal corticostriatal pathways (Vukšić, Radoš, & Kostović, 2008).

The growth of other subcortical efferent pathways—for example, long corticopontine and corticospinal axons of layer V pyramidal neurons—is less well known. Two lines of evidence suggest that subcortical efferent pathways may have reached their target structures/levels. First, microtubule maturation in cell bodies of prefrontal layer V is accelerated after 18 PCW, and expression of specific genes seems to be required for proper targeting of their axons (Chen, Rašin, Kwan, & Šestan, 2005). Second, corticospinal fibers originating in the human frontal motor region arrive at the spinal cord already at 24 PCW (Eyre, 2007; Eyre, Miller, Clowry, Conway, & Watts, 2000).

Neuronal Circuitry and Function The most differentiated neurons in the midfetal frontal cortex are polymorphic SP neurons and Cajal-Retzius cells of the MZ. The number of synapses has increased in the SP and MZ, and transient circuitry of the SP displays remarkable complexity. There is electron microscopy evidence that SP neurons are postsynaptic elements, since synapses are readily found on their proximal dendrites and cell bodies (Kostović & Rakic, 1990). The vast majority of synapses are asymmetric and therefore probably excitatory. Subplate neurons display a striking variety of neuronal morphologies (Mrzljak et al., 1988, 1990).

As shown in rhesus monkeys (Meinecke & Rakic, 1992) and carnivores (Antonini & Shatz, 1990), many SP neurons synthesize GABA, while approximately 60% of SP neurons may synthesize glutamate (Antonini & Shatz, 1990). Presynaptic elements for SP circuitry originate from monoaminergic and cholinergic afferents and other SP neurons, while thalamic axons and axons of neurons residing in the CP may represent an additional presynaptic input (Allendoerfer & Shatz, 1994). As a rule, GABAergic neurons coexpress various neuropeptides (Allendoerfer & Shatz, 1994; Delalle et al., 1997; Kostović, Štfulj-Fučić, Mrzljak, Jukić, & Delalle, 1991).

To summarize, rich presynaptic input from modulatory, nonsensory sources impinges upon a complex, transient population of neurons. Intrinsic neurons are interconnected. Thus one network communicates with external input while the other network is endogenous and very likely oscillatory.

Development of prefrontal neuronal circuitry in early preterm infants

The early preterm period is important for human development because preterm infants older than 24 PCW can survive and thus become exposed to interaction with environment. In addition, sensory stimulation is possible even in utero.

Structural Organization The prefrontal cortical regions gradually differentiate while transient fetal zones change their appearance and thickness. An initial lamination appears in the middle of the CP (figure 2.1*E*). The SP is at its developmental peak (Kostović, Lukinović, et al., 1989; Kostović & Rakic, 1990) and shows regional variations in thickness. The cytoarchitectonic immaturity of the

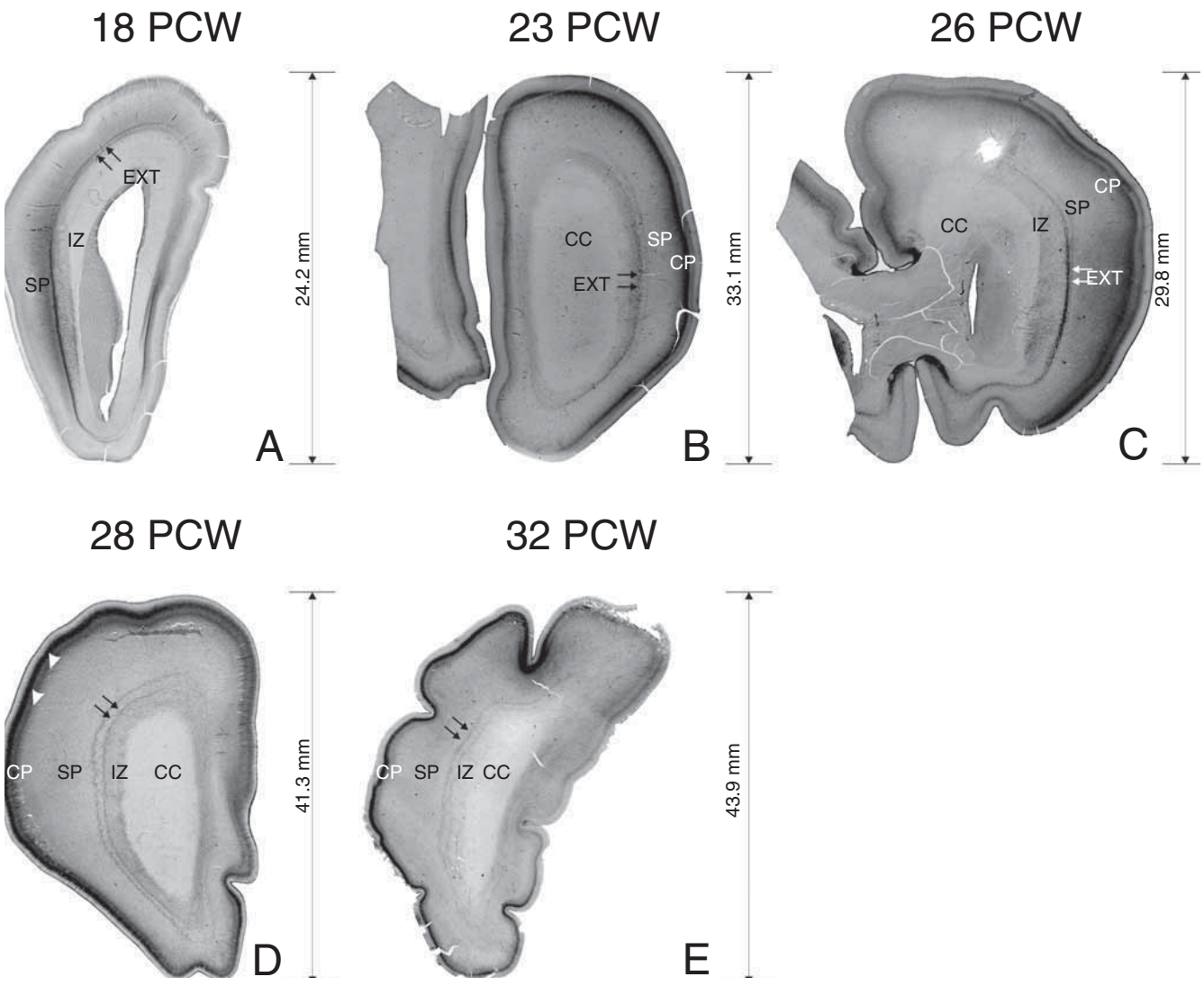


FIGURE 2.2 Laminar shifts, regional and areal differences in acetylcholinesterase (AChE) histochemical staining in the human prefrontal cortex at different developmental phases. At 18 PCW (*A*), AChE-reactive fibers originating in the basal forebrain, external capsule system (EXT), and thalamocortical/internal capsule system gradually invade the subplate. At 23 PCW (*B*), AChE-reactive fibers accumulate transiently in the superficial part of the SP. After 24 PCW, AChE-reactive fibers gradually penetrate into the cortical plate, as illustrated in 26 PCW specimen (*C*). There is a parallel decrease in staining of the subplate zone. At 28 PCW (*D*), strong AChE reactivity is present in the cortical plate (arrowheads). Gradual decrease in the overall AChE-staining is observed at 32 PCW (*E*). Double arrows point to the external capsule, that is, the deep border of the subplate zone.

early preterm prefrontal cortex is also visible in the MZ, which has several sublayers and contains well-developed but transient subpial granular layer (Kostović et al., 2004/2005).

Transient lamination of the prefrontal cortex in human preterms was well documented by acetylcholinesterase (AChE) histochemistry (Kostović, 1990). There is a transient columnar arrangement (figure 2.2*D*) of the strong AChE-staining in the prefrontal cortex, with regionally characteristic distribution (Kostović, 1990). The differences between orbital, lateral, dorsolateral, and orbitomedial cortex became more obvious and roughly correspond to the incipient gyral landmarks (Kostović, Petanjek, et al., 1992). In summary, structural data on laminar, regional, and radial organization show transient patterns of organization of prefrontal cortex in preterm infant.

NEUROGENETIC EVENTS The proliferative zones (ventricular and subventricular) in preterm infants gradually cease to produce neurons. However, the bulging of the ventricular zone, the so-called ganglionic eminence, remains thick and voluminous. According to the evidence obtained in primates (Bystron et al., 2008; Rakic, 2006) after 24 PCW this zone produces predominantly glial precursors. In our Golgi studies (Mrzljak et al., 1988, 1990) we have seen many migratory neurons in early preterm infants.

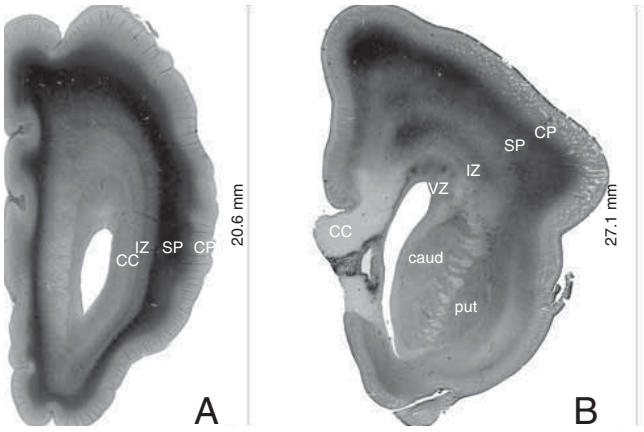


FIGURE 2.3 The subplate (SP) is characterized by an abundant extracellular matrix, as demonstrated by PAS-Alcian Blue histochemical staining (*A*). The subplate extracellular matrix also contains a high amount of axonal guidance molecules, as shown by immunostaining for fibronectin (*B*). Note gradients of extracellular matrix and fibronectin concentration within the "waiting" compartment of the subplate zone.

Therefore, it is possible that in the prefrontal pallium neurogenesis and migration last several weeks longer than in primary cortical areas.

While there is cessation of proliferative processes, both pyramidal and nonpyramidal neurons continue to differentiate. SP neurons continue to grow (Mrzljak et al., 1992) and express different neuropeptides such as NPY (Delalle et al., 1997) and somatostatin (Kostović, Štfulj-Fučić, et al., 1991), as well as GABA. The distribution of synapses changes significantly after 24 PCW: synapses begin to appear within the deep portion of the CP and bilaminar pattern of synaptic distribution gradually disappears. The number of synapses in the deep cortex (SP plus deep CP) is higher than in the superficial cortex (superficial CP plus MZ).

The predominance of deep synapses in early preterms and deep-to-superficial synaptogenesis after 28 PCW are structural factors important for generation of cortical dipole (from surface to intermediate zone) and changing surface positive potentials (Molliver, 1967). According to our studies (Kostović & Goldman-Rakic, 1983; Kostović & Judaš, 2002; Kostović & Rakic, 1984, 1990) the most intense neurogenetic process in the preterm cerebrum is growth of projection and callosal pathways.

FUNCTIONAL ORGANIZATION The main feature distinguishing preterm cortex from midfetal and late fetal cortex (figures 2.1*B*, 2.1*C*, 2.1*D*) is the presence of strong thalamic input from mediodorsal nucleus (Kostović & Goldman-Rakic, 1983). In the primary sensory areas, the thalamic input is anatomical basis for evoked potentials (for a review of literature, see Kostović & Jovanov-Milošević, 2006). After 24 PCW, thalamic afferents establish synaptic contacts with CP neurons (Molliver et al., 1973). In the somatosensory cortex of preterm infants, pain stimuli from the skin may provoke cortical response as detected by infrared monitoring (Fitzgerald, 2005). It is not known whether information from primary cortex in early preterms is conveyed further to prefrontal cortical areas. This seems to be unlikely because of the immaturity of cortico-cortical pathways. Thus the prefrontal cortex in early preterms receives predominantly nonsensory information via mediodorsal-prefrontal projection. Thalamic axons in early preterm infants make synapses with cells in both SP and CP (Kostović & Jovanov-Milošević, 2006; Kostović & Judaš, 2006, 2007).

The contact of thalamocortical axons with subplate neurons is a powerful activator of transient endogeneous cortical circuitry. The initial contact with cortical plate cells is a forerunner of extrinsic circuitry. Prolonged coexistence of these two types of circuitry, the extrinsic and transient intrinsic, is the salient feature of cortical development in humans (Kostović & Judaš, 2006, 2007). The gradual changes in organization of transient and "permanent" circuitry may form the basic framework for changing EEG (Vanhatalo & Kaila, 2006). Slow activity transients (SAT) are generated as early as 24 PCW (Tolonen, Palva, Andersson, & Vanhatalo, 2007; Vanhatalo & Kaila, 2006), transform at 30 PCW, and disappear after birth (Vanhatalo & Kaila, 2006). Excitatory extrinsic cholinergic input to the SP also originates in the basal forebrain. Subplate neurons are glutamatergic (Antonini & Shatz, 1990) and GABAergic (Meinecke & Rakic, 1992) or GABA-peptidergic neurons (Allendoerfer & Shatz, 1994). There is significant increase in number of peptidergic neurons in preterm infants (Delalle et al., 1997).

While functional relationship between thalamic axons, subplate neurons, and cortical plate neurons in primary cortical areas was clearly demonstrated in experimental studies in carnivores (Allendoerfer & Shatz, 1994) and rodents (Hanganu, Kilb, & Luhmann, 2002), the prefrontal

circuitry at this early age was not studied in monkeys, which may serve as a possible model for human circuitry. It would be of particular interest to reveal interaction of "extrinsic" and "intrinsic" circuits at synaptic and transmitter levels and correlate this with cortical potentials (Khazipov & Luhmann, 2006; Tolonen et al., 2007). In early preterm infants "intrinsic" transient circuitry of the SP may still be the main generator of spontaneous, endogeneous oscillations (Vanhatalo & Kaila, 2006) and spontaneous activity transients (Tolonen et al.). The transient circuitry of the SP in the frontal lobe may have output to caudate nucleus (Goldman-Rakic, 1981) and contribute to generation of movements in the preterm infant.

One of the most intriguing questions is when in the human prefrontal cortex GABAergic neurons became inhibitory. In experimental studies GABA was shown to act as excitatory transmitter during early phases of cortical development (Vanhatalo & Kaila, 2006) due to the immaturity of KCC transporter (Vanhatalo et al., 2005). The switch to excitatory function occurs relatively late in the rat, at the age corresponding to neonatal period in humans.

Very little is known about connections of prefrontal cortex with basal ganglia and brain stem. In the nonhuman primates, corticostriatal pathways terminate in full register with striatal cytoarchitectonic compartments at embryonic day 105 (Goldman-Rakic, 1981). In humans, striatal cell islands and AChE patches appear at 10 PCW and become maximally developed between 28 and 32 PCW (Graybiel & Ragsdale, 1978; Letinić & Kostović, 1996; Vukšić et al., 2008). The efferent pathways from the frontal cortex form the Muratoff's fascicle (Petrides & Pandya, 2007), and this fiber system seems to be the source of afferents to well-developed moduli in the human fetal putamen. In the context of transitory cortical circuitry and connections, it is interesting that the first general, variable movements appear during the preterm infant period (Hadders-Algra, 2007).

Development of neuronal circuitry of the human prefrontal cortex during the last two months of gestation

Structural Organization During the last two months of gestation, the frontal cortex and related subcortical nuclei grow rapidly. All primary gyri and sulci are recognizable at the cortical surface (figure 2.1*F*), as demonstrated by in vivo MR imaging, which, in comparison to postmortem material, provides a more accurate picture of convolutional development.

However, in classical neuroanatomical studies it was not fully appreciated that at this age laminar and areal cytoarchitectonic organization remains surprisingly immature. For example, the transient SP is still interposed between growing white matter and six-layered cortical plate. The six-layered cortical plate was described as Grundtypus by Brodmann (1909). In this six-layered pattern (figure 2.1*F*) one can distinguish individual layers, but cell sizes, morphology, and aggregation within these layers are significantly different from those in the postnatal cortex. First, layer III pyramidal neurons are small, and characteristic sublayer IIIc cannot be delineated. Second, layer IV is densely granular (figure 2.1*F*), and this granularity continues into premotor and primary motor cortex of the frontal lobe. Therefore, the criterion of layer IV granularity (agranular–dysgranular–granular) cannot be used to delineate frontal cortical areas as in the adult brain. Third, layer VI is wide and very cellular with gradual transition toward the SP. Fourth, layer I contains numerous large neurons, many of which show characteristics of Cajal-Retzius cells on Golgi preparations. The most obvious sign of immaturity is the presence of the SP, which can be demonstrated with some simple staining techniques such as PAS-Alcian Blue staining (Kostović, Judaš, et al., 2002).

However, the SP gradually diminishes in size after 32–35 PCW. The substrate for the gradual diminishment of the subplate on MR images is the decrease in the amount of extracellular matrix and other growth-related guidance molecules. The decrease in thickness and signal intensity of the SP is the hallmark of cerebral cortex on MR images in the preterm infant (Kostović, Judaš, et al., 2002). The ingrowth of long cortico-cortical pathways into the layers of the CP is another important factor in diminishment of the SP. Parallel to the decrease in the size of the SP there is development of corona radiata composed of thalamocortical and other projection fibers. The corona radiata fibers form segment III of von Monakow (1905) and continue in gyral white matter, but tight connection of the gyral white matter with cortical layers is interrupted by existence of the SP.

Neurogenetic Events Data on neuronal proliferation in nonhuman primates (Rakic, 1988, 2006) obtained in stages comparable to human development (Kostović & Rakic, 1990) indicate that we cannot expect production of new neurons in the prefrontal cortex of the late preterm infant. The notable size of proliferative ventricular-subventricular zones and ganglionic eminence are most likely related to proliferation of astroglia and oligodendroglia. The presence of areas of immature cells around the anterior horn of lateral ventricles, which are characteristic for frontal lobe, probably reflect waves of cells on the way to more superficial cortical layers. It is also very likely that some late-arriving neurons still migrate because the migratory radial route is longer in prefrontal cortex than in other cortical regions.

Growth of dendrites is a significant event in preterm infant (Mrzljak et al., 1988, 1990). This phase of differentiation is influenced by glutamatergic thalamic afferents (Mrzljak et al., 1988, 1990) and is described as a second phase of

dendritic differentiation in the prefrontal cortex (Petanjek et al., 2008). Parallel to this event there is initial development of characteristic dendritic orientation in nonpyramidal, local circuitry neurons (Mrzljak et al., 1988; Kostović, Judaš, & Petanjek, 2008). The most robust neurogenetic event in the preterm infant is growth of long cortico-cortical pathways. Callosal fibers are in the phase of growth around ventricles (Kostović, Judaš, et al., 2002) and show significant increase in number, described as exuberance (Chalupa & Killackey, 1989; Innocenti & Price, 2005).

There is no reliable chemical marker for direct demonstration of long cortico-cortical associative fibers. The DT tractography demonstrated preliminary results without identifying source(s) of associative pathways described in adult prefrontal cortex (Petrides & Pandya, 2007). The conventional MRI shows periventricular zone rich in extracellular matrix (Judaš et al., 2005), which corresponds to axonal strata containing long associative fascicles such as fronto-occipital and superior longitudinal medial fasciculus described by von Monakow (1905) in developing fetal and child brain. In addition, fiber bundles that contain efferent associative pathways from the rostral prefrontal cortex (Petrides & Pandya, 2007), such as external capsule and cingulate fasciculus, are well developed in the preterm infant.

Functional Organization There is further maturation of evoked and event-related potentials: the gradual appearance of synchronous EEG activity and slow activity transients become less frequent. Analysis of general movements that depend on transient circuitry (Hadders-Algra, 2007) shows characteristic "writhing" general movements (Hadders-Algra, 2007). Because of the significant synaptogenesis in superficial layers of the cortical plate, cortical dipole switches from predominantly deep to predominantly superficial generation of electrical activity (Molliver, 1967). Accordingly, preterm infants show surface-negative cortical potentials that change to surface-positive potentials during the early postnatal period.

As stated for early preterm infants, the significance of sensory input for maturation of the prefrontal cortex is not known. In experiments performed in monkey fetuses delivered 3 months before the term and exposed to normal light, intense visual stimulation did not affect the rate of synaptic production (Bourgeois, Jastreboff, & Rakic, 1989). Accordingly, one can assume that fetal afferent input to prefrontal cortex before the birth will not affect the rate of synaptic production.

Neonatal period

Structural Organization The development of gyri and sulci continues with formation of tertiary convolutional pattern (Armstrong, Schleicher, Omran, Curtis, & Zilles, 1995). The prefrontal cortex is cytoarchitectonically immature, and the main reason for this immaturity is that granular layer IV, a main recipient of thalamocortical input, remains well developed in all frontal areas (i.e., in both prospective granular and prospective dysgranular areas). In layer I (the former fetal marginal zone) there is concentration of neurons, and well-developed Cajal-Retzius cells are still present (Krmpotić-Nemanić, Kostović, Vidić, Nemanić, & Kostović-Knežević, 1987).

The newborn cortex has immature laminar organization with prominent layer IV, dense layer II, absence of magnopyramidality of layer IIIc, and wide transition between layer VI and the white matter—cytoarchitectonic remnants of the subplate zone (figure 2.4). The precise delineation of dysgranular areas (Broca's area 44, posterior orbital cortex) from prefrontal granular areas is not possible on the basis of cytoarchitectonic criteria (Judaš & Cepanec, 2007).

Neurogenetic Events The neonatal period is dominated by dendritic differentiation (figure 2.5), growth and reorganization of callosal fibers, and growth of short cortico-cortical fibers. Dendritic development is evident in all classes of neurons (Mrzljak et al., 1990). At birth, layer V pyramidal neurons have larger and more complex dendritic trees than those of layer IIIC (Petanjek et al., 2008). Pyramidal neurons of layers V and III display the first phase of spinogenesis and intense growth of basal dendrites. Interneurons of layer IV develop characteristic phenotypes and can be distinguished. The supragranular layers contain double-bouquet neurons, basket neurons, and multipolar and bitufted nonpyramidal neurons (Mrzljak et al., 1990).

It is interesting that subplate neurons show continuous dendritic growth (Mrzljak et al., 1988, 1990, 1992). The growth of short cortico-cortical pathways was not directly demonstrated in the human prefrontal cortex, but in the human visual cortex DiI tracing demonstrated growth of short cortical pathways during corresponding phases of development (Burkhalter, 1993). Synaptogenesis is quantitatively significant neurogenetic event in the neonatal cortex, and there is rapid increase in the formation of synapses (Huttenlocher & Dabholkar, 1997; Kostović, Judaš, Petanjek, & Šimić, 1995).

Functional Organization During the neonatal period, thalamocortical circuitry is established and there is a substrate for interaction with other brain regions and sensory environment. This process is very important for shaping of fine connectivity within the cortical columns (Penn & Shatz, 1999). During the neonatal period, the anatomical substrate of electrical activity is rapidly changing because of exponential increase in the number of synapses, deep synaptic activation, retraction of exuberant callosal axons (Chalupa & Killackey, 1989; Innocenti & Price, 2005), and ingrowth of

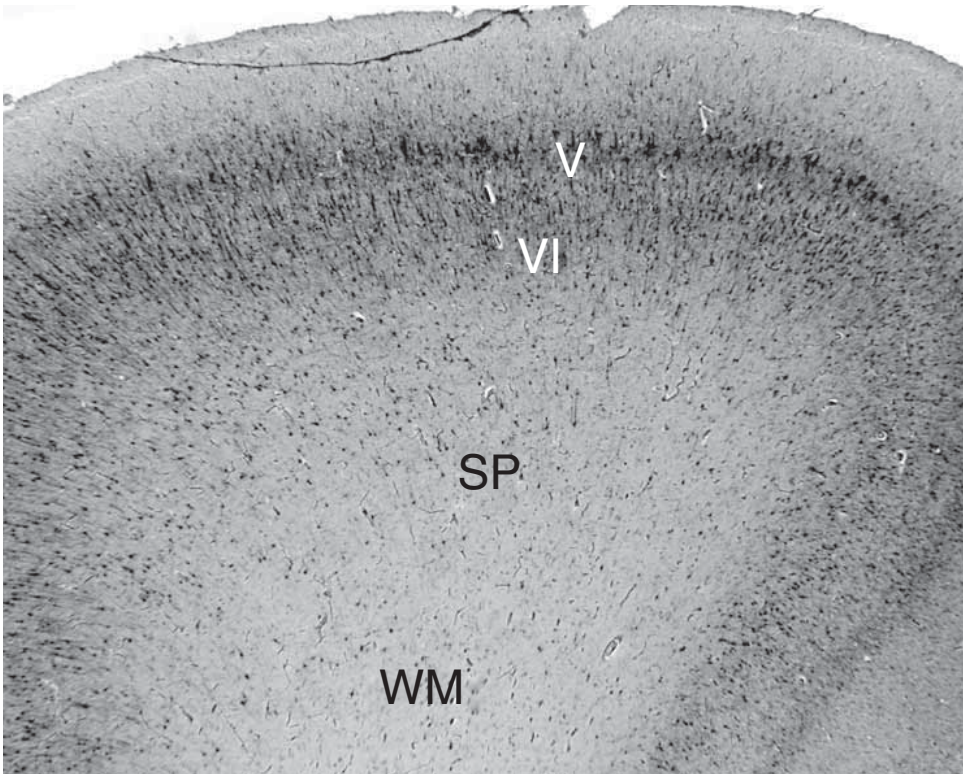


FIGURE 2.4 Transient subplate neurons are present in the postnatal human cerebral cortex, as revealed by MAP2-immunohistochemistry of the middle frontal gyrus. MAP2-positive neurons are seen extending from deep cortical layer VI through the SP into the core of the gyral white matter (WM).

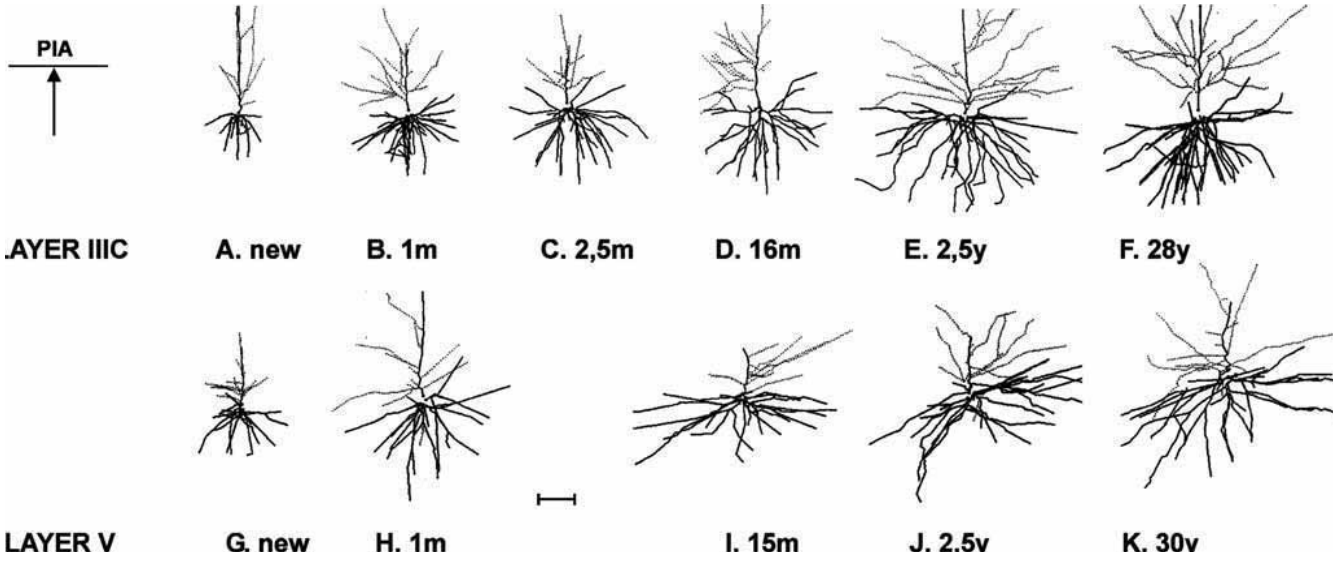


FIGURE 2.5 Development of dendritic arborizations of layers III and V pyramidal neurons, as revealed by computerized Neurolucida reconstructions. (After Petanjek et al., 2008, with permission of Oxford University Press.)

cortico-cortical fibers. As a consequence the cortical dipole changes and the cortical surface response becames predominantly positive (Novak, Kurtzberg, Kreuzer, & Vaughan, 1989). The EEG shows more continuous activity (Dreyfus-Brisac, 1979). Slow activity transients may be still present in the full-term newborn (Vanhatalo et al., 2005; Vanhatalo & Kaila, 2006). This phenomenon, together with prolonged existence of transient circuitry and immaturity of cortico-cortical connectivity, points to the overall functional immaturity of the cortex. Supragranular local circuitry neurons have nonmyelinated axons and immature dendritic arborizations (Mrzljak et al., 1990). This immaturity of neuronal circuitry correlates well with immaturity of the psychological basis for recruiting and sustaining attention to a stimulus that is guided primarily by its physical features (Kagan & Baird, 2004).

Infancy

Structural Organization The laminar and areal differentiation shows substantial differences between early (2–3 months) and late infancy (7–12 months). In the early infant cortex (figure 2.1*G*) granular layer IV is present throughout polar, orbital, medial, and dorsolateral prefrontal cortex (Kostović, 1990), as well as in the fronto-opercular region (Judaš & Cepanec, 2007). In the late infant cortex (figure 2.1*H*) dysgranular areas of posterior orbital and fronto-opercular cerebral cortex may be delineated with greater confidence. Accordingly, Amunts, Schleicher, Ditterich, and Zilles (2003) found left-right asymmetry in Broca's region (Brodmann's areas 44 and 45) in 1-year-old infants. The other difference between early and late infancy is the presence of transient subplate. In early infants (2–3 months) the subplate is visible as transitional layer situated between layer VI and white matter core of frontal gyri (Kostović, Petanjek, et al., 1992). The presence of transient subplate elements obscures measurements of cortical thickness on MRI scans (Giedd et al., 1999; Sowell et al., 2004). Further structural differences between early and late infant brain are related to the maturation of pyramidal neurons of layers V and III (figure 2.5). Progressive differentiation of pyramidal neurons continues in infancy in both supragranular and infragranular layers (Mrzljak et al., 1988, 1990, 1992; Petanjek et al., 2008). During the first postnatal month, the extent of dendritic differentiation is similar in pyramidal neurons of layers III and V. However, after the third month, layer IIIC pyramidal cells display a seemingly "dormant" period with no significant overall dendritic growth from 3 to 16 months (Petanjek et al., 2008).

Neurogenetic Events During infancy there is continuing growth of short cortico-cortical and intracortical connections, retraction of exuberant callosal projections, development of dendrites of both pyramidal and inhibitory neurons, explosive synaptogenesis, naturally occurring cell death, and incipient myelogenesis (figure 2.6). The growth of short cortico-cortical connections continues at least until the 6th postnatal month, but may continue even after this period (Burkhalter, 1993). The retraction of exuberant axons occurs before the 6th postnatal month (Innocenti & Price, 2005), which is in agreement with experimental results in monkey (Chalupa & Killackey, 1989; LaMantia & Rakic, 1990). In the newborn monkey there are 3.5 times more callosal axons than in the adult (LaMantia & Rakic, 1990). Synaptic density increases rapidly during the first postnatal year (Huttenlocher & Dabholkar, 1997). Similar results were obtained in counting of spines in the human prefrontal cortex (Kostović, Petanjek, et al., 1992). After the 7th postnatal month, the growth of cortico-cortical pathways subsides, and predominant neurogenetic events are synaptogenesis, dendritogenesis, and myelogenesis.

Functional Organization It is not known how the prefrontal cortex participates in behavior of infants during the first postnatal months. The first novel phase of behavior with clear involvement of the prefrontal cortex is goal-directed behavior (Kostović, Judaš, et al., 1995) demonstrated in the human version of the delayed response test (Diamond & Goldman-Rakic, 1989) where engagement of prefrontal cortex may be demonstrated between 7 and 10 months of postnatal age. It is obvious that this complex behavior depends on a new level of cortico-cortical and intrinsic cortical circuitry organization. The construction and shaping of this circuitry lasts almost a year and may be divided into early (2–3 months) and late phases (4th month onward).

Both phases are characterized by transitional patterns of cortical organization. These phases correspond to two transitions described for psychological development during the first year (Kagan & Baird, 2004). During the early infancy phase (2–3 postnatal months) there is a rapid increase in number of synapses and growth of short cortico-cortical pathways (Burkhalter, 1993; Huttenlocher & Dabholkar, 1997; Kostović, Petanjek, et al., 1992). In addition, there is a rapid dendritic development of pyramidal neurons (Petanjek et al., 2008).

During the first postnatal months some subplate neurons continue to grow and maintain transient circuitry. Some authors consider transient subplate circuitry as a framework for general movements during the neonatal period and early infancy (Hadders-Algra, 2007). However, participation in other transient phenomena cannot be excluded. It is not known whether subplate resolution is related to retraction of callosal and long associative fibers that entered this zone but never reached cortical plate layers. It is also not known what is the contribution of these pathways to transient cortical

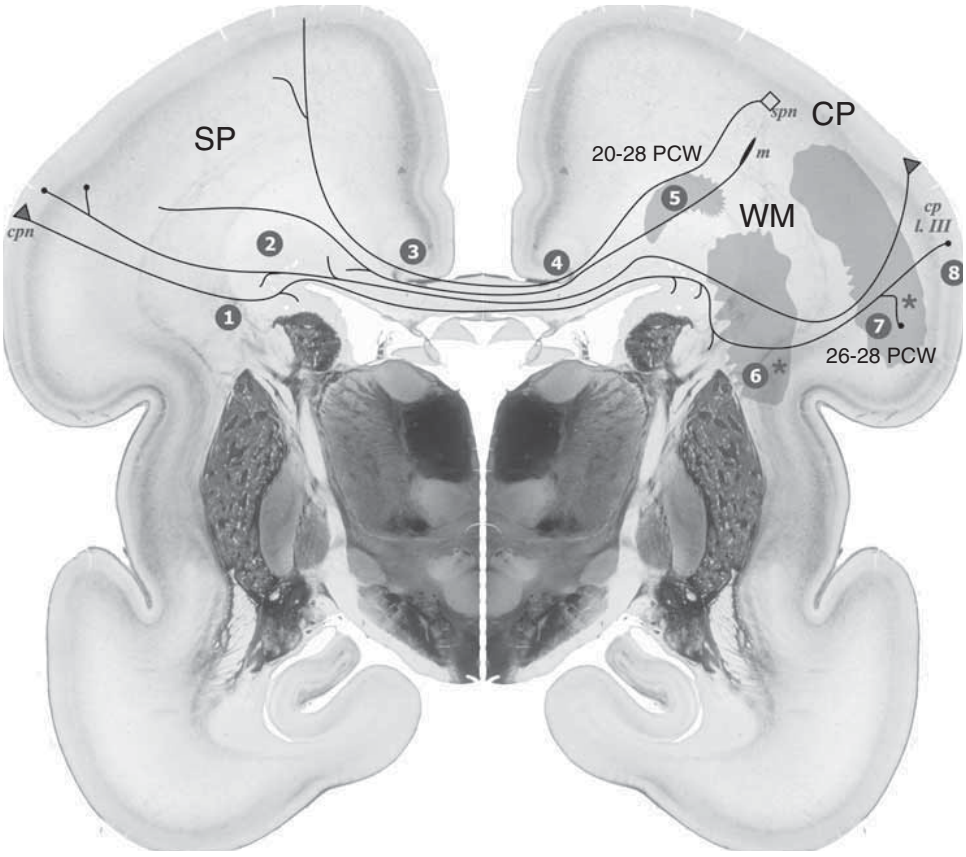


FIGURE 2.6 Growth of corpus callosum fibers through complex guidance zones and decision points. Numbers denote eight sequential decision points in the ipsilateral hemisphere, at the hemispheric midline, and in the opposite hemisphere. Asterisks indicate areas of intermingling with thalamocortical fibers. Abbreviations: *cp*, cortical plate; *cpn*, cortical plate neuron; *l.III*, developing layer III; *m*, migrating neuron; SP, subplate; *spn*, subplate neuron.

circuitry. Although the pruning of exuberant callosal connections in the human brain occurs during the first six months of postnatal life, the moderate reduction in size of the corpus callosum was noted already in preterm infants (Innocenti & Price, 2005).

We proposed that the subplate exists as a cytoarchitectonic entity during growth of long cortico-cortical pathways and that some subplate circuitry elements exist also during the late transitional phase of growth of short cortico-cortical connectivity pathways (Kostović & Rakic, 1990; Kostović, 1990; Kostović, Petanjek, et al., 1992). This late phase of transient circuitry is resolved between 7 and 10 months of postnatal life when cortico-cortical connectivity is established and when prefrontal cortex exhibits first executive functions (Diamond & Goldman-Rakic, 1989; Fuster, 2002; Goldman-Rakic, 1987). This finding suggests that the disappearance of transient circuitry (which existed in parallel with permanent circuitry elements since 24 PCW) should precede the onset of goal-directed behavior and ability to retrieve schemas from past events that are no longer in the perceptual field. We describe this new cognitive phase, which develops after the 7th postnatal month, as environmentally driven and as marking the onset of characteristic executive function with resolution of spatiotemporal relationships.

As stated earlier, the maturation of layer III neurons displays a "dormant" period between the 3rd and 15th postnatal months (Petanjek et al., 2008). According to these data, the differentiation of associative neurons of layer IIIC speeds up after the 15th month, during the period of intensified social interaction.

The development of subplate peptidergic neurons follows a similar time schedule. In the newborn, NPY and somatostatin neurons are important constituents of the subplate (Delalle et al., 1997; Kostović, Štfulj-Fučić, et al., 1991). However, by 6 months, their number decreases, and this low number is maintained during childhood (Delalle et al., 1997).

Modulatory pathways from basal forebrain and brain stem change their transient fetal pattern during early infancy (Kostović, Škavić, & Strinović, 1988; Brown, Crane, & Goldman, 1979). However, important developmental shifts in their density and distribution occur during childhood and adolescence (Brown et al., 1979).

There is a correlation between synaptogenesis (Huttenlocher & Dabholkar, 1997), formation of dendritic spines

(Kostović, Petanjek, et al., 1992; Petanjek et al., 2008), and elaboration of dendrites in layer V and III pyramidal neurons (Petanjek et al., 2008). This indicates that in the early cognitive phase the main event in the formation of prefrontal neuronal circuitry is differentiation of pyramidal neurons and their synaptic connections.

In conclusion, the circuitry of the early cognitive phase (7–12 months) is characterized by finalization of growth of short cortico-cortical pathways, rapid growth of layer V pyramidal dendrites (Petanjek et al., 2008), initial differentiation of inhibitory neurons, especially double-bouquet interneurons (Mrzljak et al., 1990), and the disappearance of the subplate.

Early childhood: The second year

Structural Organization During the second year, areal specification of the prefrontal cortex is in its final phase but not yet completed. Two main cytoarchitectonic features of the prefrontal cortex, magnopyramidality of layer IIIC and granularity of layer IV, help to distinguish dorsolateral, dorsolateral-medial, polar, orbital, medial, and precingulate areas. At the beginning of the second year layer IIIC pyramidal neurons are still in the "dormant" period characteristic of infancy (Petanjek et al., 2008). However, the maturation of their cytoskeleton advanced significantly when compared with findings in early infancy (Jovanov-Milošević, Petrović, Judaš, & Kostović, 2008).

At the end of the second year (figure 2.1*I*), layer IIIC pyramidal neurons become larger than those in layer V and dorsolateral prefrontal areas acquire magnopyramidal appearance (figure 2.5). Around the second birthday, dendrites of layer IIIC pyramidal neurons enter the second phase of significant elongation. In contrast to the biphasic pattern of growth of layer III pyramidal neurons, pyramidal neurons of layer V projection neurons show a linear pattern of development confined to infancy.

Neurogenetic Events Three main neurogenetic events characterize early childhood: (1) synaptogenesis, (2) dendritic differentiation of both pyramidal and nonpyramidal local circuitry neurons, and (3) myelogenesis (figure 2.7). The formation of synapses is progressive and more intense during the second year of life (Huttenlocher & Dabholkar, 1997; Kostović, Petanjek, et al., 1992; Petanjek et al., 2008). The number of dendritic spines reaches its maximum at 2.5 years of life (Kostović, Petanjek, et al., 1992; Petanjek et al., 2008). The differentiation of dendrites in layer IIIC neurons shows a similar trend: after a yearlong "dormant" period, pyramidal neurons display a second period of dendritic growth at the end of the second year (Petanjek et al., 2008).

Based on quantitative parameters described for human prefrontal cortex (discussed earlier), the differentiation of dendritic trees and growth of fine, distal dendrites is another major histogenetic event in both pyramidal and nonpyramidal neurons during the second year of life. It seems likely that these new postsynaptic sites participate in establishment of late-maturing intracortical circuitry.

The prefrontal cortex of the rhesus monkey contains 13 different classes of neurons that constrain the propagation of pyramidal cell excitation in different projections including local circuits (Lewis, Hashimoto, & Volk, 2005; Lund & Lewis, 1993).

The process of myelinization involves also predominantly long cortico-cortical pathways. Some long cortico-cortical pathways such as corpus callosum mature as late as 10 years of life (Yakovlev & Lecours, 1967). The most interesting is myelinization of intracortical fibers that form the final, fifth segment of the cortical white matter (von Monakow, 1905) described as radii. The intracortical portion of myelinated fibers cannot yet be demonstrated with conventional MR imaging and was described in classical myeloarchitectonic studies (von Monakow, 1905; Yakovlev & Lecours, 1967).

Functional Organization During the second year, new cognitive functions gradually develop (Kagan & Baird, 2004). These new competences are largely integrated in the prefrontal cortex: production of meaningful speech, the capacity to infer selected mental states in others (theory of mind), the first signs of conscious awareness, and feeling of the self (Kagan & Baird, 2004; Werker & Vouloumanos, 2001).

The maturational shift that is most likely to underlie these complex environment-brain interactions is related to intrinsic cortical circuitry. First, it concerns the level of differentiation of layer IIIC pyramidal neurons achieved around the second birthday. Second, it is related to differentiation of GABAergic inhibitory interneurons: double-bouquet, calretinin-containing neurons; chandelier, parvalbumin-containing neurons; and wide-arbor basket, parvalbumin-containing neurons. Activation of interneurons serves to constrain the propagation of pyramidal cell excitation in local and long-range as well as intrinsic and associative projections.

The increased number of postsynaptic (Mrzljak et al., 1990; Petanjek et al., 2008) and presynaptic intrinsic elements (Anderson, Classey, Conde, Lund, & Lewis, 1995) between 2 and 2.5 years corresponds to peak density of dendritic spines and synapses observed in this period. This correspondence was shown in studies of synaptogenesis (Huttenlocher & Dabholkar, 1997) and spinogenesis (Kostović, Petanjek, et al., 1992; Kostović, Judaš, & Petanjek, 2008; Petanjek et al., 2008). It seems that this second cognitive phase requires the presence of supernumerary synapses in order to selectively stabilize functional circuitries through interaction with different environmental cues.

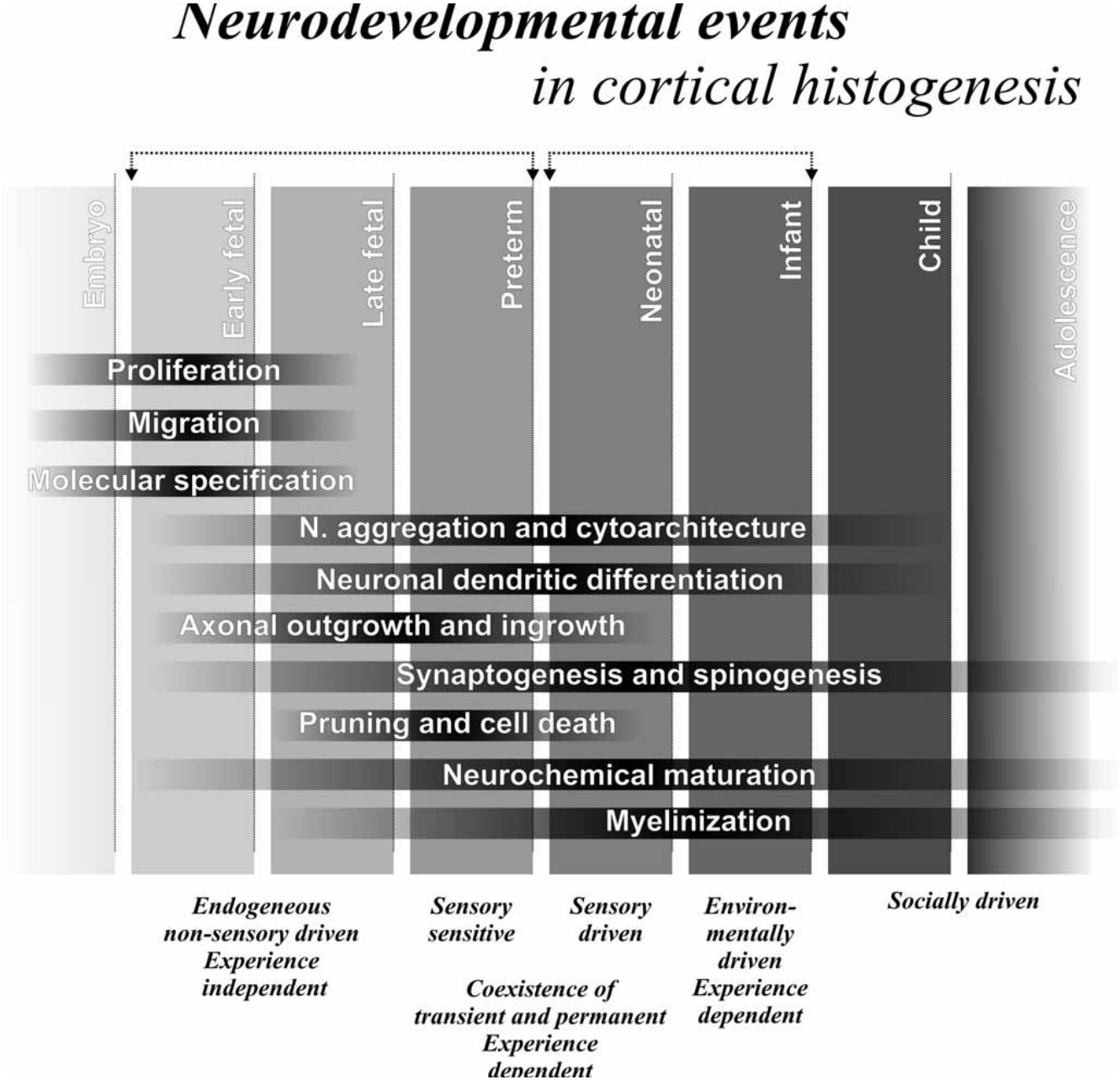


FIGURE 2.7 Simplified graphic representation of timing and intensity of neurogenetic events and different types of interaction with environment. Note the differences in intensity of neurogenetic events during prenatal and postnatal development.

Overview of the developmental phases of prefrontal circuitry during the first 2 years of life

The data presented systematically in previous paragraphs clearly suggest that development of prefrontal circuitry begins during the early fetal life, in parallel with development of regions destined for motor and sensory functions (Kostović & Rakic, 1990).

These early circuits are oscillatory, endogeneous, and very likely based on nonsynaptic contacts because we did not see synapses before the formation of the cortical plate. The first synaptic circuitry involves postsynaptic elements situated in characteristic bilaminar distribution above and below the cortical plate (figure 2.8*A*). This phase is endogeneous and nonsensory driven. The crucial event in the development of fetal circuitry occurs after 13 PCW when cortical plate loosens and new prominent subplate zone develops. This typical fetal circuitry consists of (1) postsynaptic elements: well-differentiated GABAergic and glutamatergic subplate neurons, Cajal-Retzius cells of the marginal zone, and basal and apical arborizations of quite immature pyramidal neurons; (2) presynaptic elements: short and long axons of

glutamatergic and GABAergic subplate neurons; and (3) modulatory elements: afferents from monoaminergic brain stem neurons and cholinergic neurons from rostral cell groups of the basal forebrain.

Thalamic afferents from mediodorsal nucleus project heavily to the subplate zone. The phase of typical fetal lamination lasts until 21–23 PCW (figure 2.8*B*) when thalamocortical fibers accumulate below the cortical plate. This phase we describe as endogeneous due to the fact that frontal cortex cannot receive extrinsic inputs because primary cortical regions are not yet driven by sensory inputs.

In contrast to the rich synaptic and transmitter organization of the subplate zone with its GABA and glutamatergic cell networks (Meinecke & Rakic, 1992; Allendoerfer & Shatz, 1994) triggered by cholinergic and monoaminergic input (Dupont et al., 2006; Molliver, 1967; Zecevic & Verney, 1995), the cortical plate is immature and does not contain synaptic contacts (Molliver et al., 1973). The pyramidal neurons show the initial phase of maturation (Mrzljak et al., 1988, 1990, 1992) by both qualitative and quantitative criteria. Nonpyramidal neurons of the cortical plate are immature (Mrzljak et al., 1988) because they are born after 18 PCW.

The next phase of circuitry development occurs at age corresponding to preterm birth after 24 PCW (figure 2.8*C*) and is characterized by penetration of CP by thalamocortical fibers from mediodorsal nucleus (Kostović & Goldman-Rakic, 1983). It is not known what kind of information is conveyed through mediodorsal nucleus, but activation of cortex is expected in this period.

In late preterm infants cortico-cortical, commissural fibers of the corpus callosum and long associative pathways establish the first cortico-cortical network. These connections also display transient exuberancy like corpus callosum (Innocenti & Price, 2005). However, many commissural axons and associative pathways do not necessarily make synapses at this stage. The overall number of synapses is relatively small in comparison to postnatal stages (Bourgeois et al., 1994; Huttenlocher & Dabholkar, 1997; Kostović, Petanjek, et al., 1992; Petanjek et al., 2008; Rakic, 2006).

The synchronous appearance of EEG suggests a certain degree of cortical integration (Albrieux et al., 2004; Khazipov & Luhmann, 2006; Kostović, Judaš, et al., 1995; Tolonen et al., 2007; Vanhatalo & Kaila, 2006; Vanhatalo et al., 2005). Efferent pathways to subcortical targets (Eyre, 2007; Eyre et al., 2000) and striatum (Goldman-Rakic, 1981; Kostović, 1984; Vukšić et al., 2008) set up the stage for later involvement in executive functions.

The circuitry in the preterm period is different from neonatal because long associative pathways are still growing, elaborate transient subplate circuitry coexists with permanent connectivity of cortical layers (thalamocortical circuitry), and areal differentiation is in its initial stage. During the neonatal period cortex becomes accessible to environmental sensory inputs. Accordingly, thalamocortical circuitry drives cortical activity in sensory cortical areas.

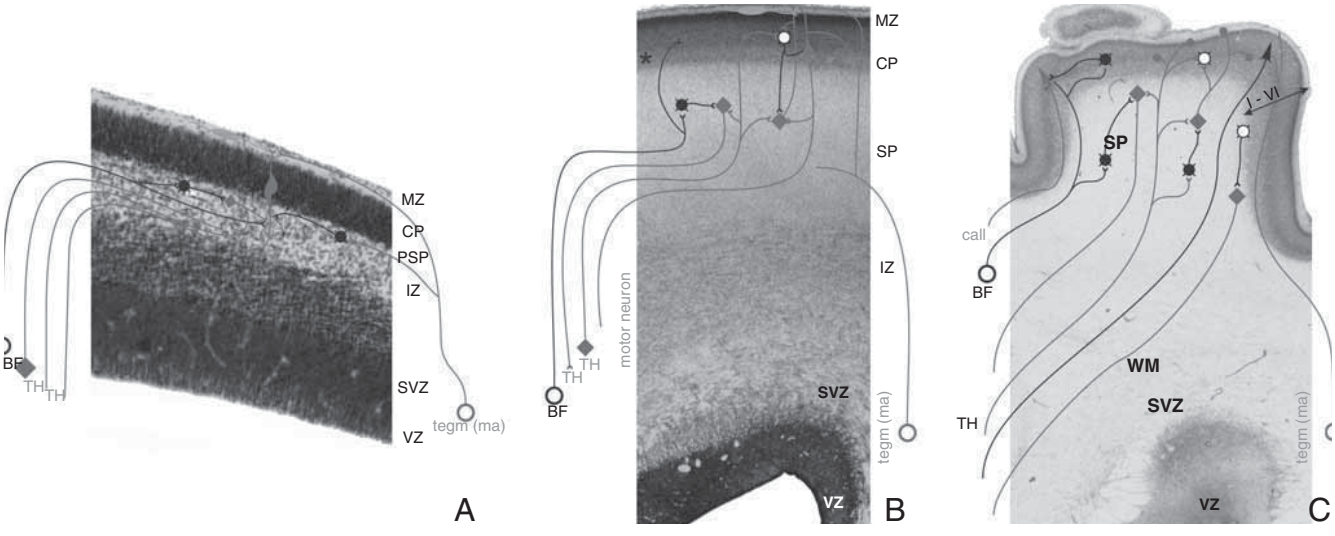


FIGURE 2.8 The reconstruction of neuronal connections in the human fetal frontal cortex. At 10 PCW (*A*), early modulatory afferents are present in transient fetal subplate (SP) and marginal zone (MZ). In the midfetal cortex at 24 PCW (*B*), thalamocortical afferents and SP neurons are dominant elements of deep transient cortical circuitry. The relocation of afferents from SP into the CP with subsequent synaptogenesis are illustrated for early preterm infant cortex at 34 PCW (*C*). (After Kostović & Judaš, 2007, with permission of Elsevier Science Publishers.) Abbreviations: BF, basal forebrain; TH, thalamus; tegm (ma), tegmental monoaminergic afferents. Black squares depict GABAergic subplate neurons, and white circles depict GABAergic neurons of the cortical plate. Glutamatergic neurons are depicted as diamond shapes.

The redistribution of thalamocortical terminals during the early preterm period, the redistribution of cortico-cortical fibers during the perinatal (late preterm and neonatal) period, and gradual changes in the transient pattern of cortical organization all suggest that there is a profound reorganization of developing prefrontal circuitry. We described this series of perinatal neurogenetic events as the "first" reorganization (Kostović, 1990), whereas early postnatal changes of cortico-cortical systems were described as the "second" reorganization (Kostović, 1990). After each reorganization, one can recognize a new phase of circuitry development, with more advanced level of functional organization (Kostović, 1990).

However, the main prefrontal circuitry elements, pyramidal neurons of layer III, remain small and immature during the neonatal period, and growth of their dendrites lags behind those of layer V pyramidal cells. Thus the sensory-driven neonatal phase observed in development of primary cortical areas (Kostović & Judaš, 2006, 2007) may have a limited influence on functions of frontal cortex because of the immaturity of cortico-cortical connections. This period is followed by two transitional phases that prepare the prefrontal cortex for its executive and goal-directed functions. These phases are characterized by existence of transient elements in layer I and subplate, the rapid increase in synaptogenesis, and the differentiation of cortical interneurons. Between 2 and 6 postnatal months, infants increasingly respond to environmental stimuli. The first overtly cognitive phase occurs between 7 and 10 months, after the disappearance of transient fetal circuitry elements, significant increase in number of synapses, and the completion of growth of short cortico-cortical fibers.

The "cognitive" prefrontal circuitry shows further quantitative and qualitative changes during the second postnatal year. Dendrites of layer IIIC pyramidal neurons enter the "second" growth spurt at the end of the second year, after a protracted dormant period. The number of synapses also continues to increase. These processes reach their peak around the second birthday, when they contribute to the emergence of speech and verbal communication.

ACKNOWLEDGMENTS This work was supported by Croatian Ministry of Science, Education, and Sport grants No. 108-1081870-1876 (to I.K.) and No. 108-1081870-1878 (to M. J.). We gratefully acknowledge the excellent technical assistance of Pero Hrabač in preparation of figures.

REFERENCES

ALBRIEUX, M., PLATEL, J. C., DUPUIS, A., VILLAZ, M., & MOODY, W. J. (2004). Early expression of sodium channel transcripts and sodium current by Cajal-Retzius cells in the preplate of the embryonic mouse neocortex. *J. Neurosci.*, *24*, 1719–1725.

ALLENDOERFER, K. L., & SHATZ, C. J. (1994). The subplate, a transient neocortical structure: Its role in the development of connections between thalamus and cortex. *Annu. Rev. Neurosci.*, *17*, 185–218.

AMUNTS, K., SCHLEICHER, A., DITTERICH, A., & ZILLES, K. (2003). Broca's region: Cytoarchitectonic asymmetry and developmental changes. *J. Comp. Neurol.*, *465*, 72–89.

ANDERSON, S. A., CLASSEY, J. D., CONDE, F., LUND, J. S., & LEWIS, D. A. (1995). Synchronous development of pyramidal neuron dendritic spines and parvalbumin-immunoreactive chandelier neuron axon terminals in layer III of monkey prefrontal cortex. *Neuroscience*, *67*, 7–22.

ANTONINI, A., & SHATZ, C. J. (1990). Relation between putative transmitter phenotypes and connectivity of subplate neurons during cerebral cortical development. *Eur. J. Neurosci.*, *2*, 744–761.

ARMSTRONG, E., SCHLEICHER, A., OMRAN, H., CURTIS, M., & ZILLES, K. (1995). The ontogeny of human gyrification. *Cereb. Cortex*, *5*, 56–63.

BOURGEOIS, J. P., GOLDMAN-RAKIC, P. S., & RAKIC, P. (1994). Synaptogenesis in the prefrontal cortex of rhesus monkeys. *Cereb. Cortex*, *4*, 78–96.

BOURGEOIS, J. P., JASTREBOFF, P. J., & RAKIC, P. (1989). Synaptogenesis in visual cortex of normal and preterm monkeys: Evidence for intrinsic regulation of synaptic overproduction. *Proc. Natl. Acad. Sci. USA*, *86*, 4297–4301.

BRODMANN, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: Johann Ambrosius Barth.

BROWN, R. M., CRANE, A., & GOLDMAN, P. S. (1979). Regional distribution of monoamines in the cerebral cortex and subcortical structures of the rhesus monkey: Concentrations and in vivo synthesis rates. *Brain Res.*, *168*, 133–150.

BURKHALTER, A. (1993). Development of forward and feedback connections between areas V1 and V2 of human visual cortex. *Cereb. Cortex*, *3*, 476–487.

BYSTRON, I., BLAKEMORE, C., & RAKIC, P. (2008). Development of the human cerebral cortex: Boulder Committee revisited. *Nat. Rev. Neurosci.*, *9*, 110–122.

BYSTRON, I., RAKIC, P., MOLNAR, Z., & BLAKEMORE, C. (2006). The first neurons of the human cerebral cortex. *Nat. Neurosci.*, *9*, 880–886.

CASEY, B. J., GIEDD, J. N., & THOMAS, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biol. Psychol.*, *54*, 241–257.

CASEY, B. J., TOTTENHAM, N., LISTON, C., & DURSTON, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends Cogn. Sci.*, *9*, 104–110.

CHALUPA, L., & KILLACKEY, H. P. (1989). Process elimination underlies ontogenetic change in the distribution of callosal projection neurons in the postcentral gyrus of the fetal rhesus monkey. *Proc. Natl. Acad. Sci. USA*, *86*, 1076–1079.

CHEN, J. G., RAŠIN, M. R., KWAN, K. Q., & ŠESTAN, N. (2005). Zfp312 is required for subcortical axonal projections and dendritic morphology of deep-layer pyramidal neurons of the cerebral cortex. *Proc. Natl. Acad. Sci. USA*, *102*, 17792–17797.

CHUGANI, H. T., PHELPS, M. E., & MAZZIOTTA, J. C. (1987). Positron emission tomography study of human brain functional development. *Ann. Neurol.*, *22*, 487–497.

DELALLE, I., EVERS, P., KOSTOVIĆ, I., & UYLINGS, H. B. M. (1997). Laminar distribution of neuropeptide Y-immunoreactive neurons in human prefrontal cortex during development. *J. Comp. Neurol.*, *379*, 515–522.

Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and infant rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Exp. Brain Res.*, *74*, 24–40.

Dreyfus-Brisac, C. (1979). Ontogenesis of brain bioelectrical activity and sleep organization in neonates and infants. In F. Falkner and J. M. Tanner (Eds.), *Human Growth*, Vol. 3: *Neurobiology and Nutrition* (pp. 157–182). London: Bailliere-Tindall.

Dupont, E., Hanganu, I. L., Kilb, W., Hirsch, S., & Luhmann, H. J. (2006). Rapid developmental switch in the mechanisms driving early cortical columnar networks. *Nature*, *439*, 79–83.

Eyre, J. A. (2007). Corticospinal tract development and its plasticity after perinatal injury. *Neurosci. Biobehav. Rev.*, *31*, 1136–1149.

Eyre, J. A., Miller, S., Clowry, G. J., Conway, E. A., & Watts, C. (2000). Functional corticospinal projections are established prenatally in the human foetus permitting involvement in the development of spinal motor centres. *Brain*, *123*, 51–64.

Fitzgerald, M. (2005). The development of nociceptive circuits. *Nat. Rev. Neurosci.*, *6*, 507–520.

Friauf, E., & Shatz, C. J. (1991). Changing patterns of synaptic input to subplate and cortical plate during development of visual cortex. *J. Neurophysiol.*, *66*, 2059–2071.

Fuster, J. M. (2002). Frontal lobe and cognitive development. *J. Neurocytol.*, *31*, 373–385.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nat. Neurosci.*, *2*, 861–863.

Goldman-Rakic, P. S. (1981). Prenatal formation of cortical input and development of cytoarchitectonic compartments in the neostriatum of the rhesus monkey. *J. Neurosci.*, *7*, 721–735.

Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive function. *Child Dev.*, *58*, 642–691.

Graybiel, A. M., & Ragsdale, J. C. W. (1978). Histochemically distinct compartments in the striatum of human, monkey and cat demonstrated by acetylcholinesterase staining. *Proc. Natl. Acad. Sci. USA*, *75*, 5723–5726.

Grove, E. A., & Fukuchi-Shimogori, T. (2003). Generating the cerebral cortical area map. *Annu. Rev. Neurosci.*, *26*, 355–380.

Hadders-Algra, M. (2007). Putative neural substrate of normal and abnormal general movements. *Neurosci. Biobehav. Rev.*, *31*, 1181–1190.

Hammock, E. A. D., & Levitt, P. (2006). The discipline of neurobehavioral development: The emerging interface of processes that build circuits and skills. *Hum. Dev.*, *49*, 294–308.

Hanganu, I., Kilb, L., & Luhmann, H. J. (2002). Functional synaptic projections onto subplate neurons in neonatal rat somatosensory cortex. *J. Neurosci.*, *22*, 7165–7176.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.*, *387*, 167–178.

Innocenti, G. M., & Price, D. J. (2005). Exuberance in the development of cortical networks. *Nat. Rev. Neurosci.*, *6*, 955–965.

Jovanov-Milošević, N., Petrović, D., Judaš, M., & Kostović, I. (2008, July 12–16). *Transition from regional to areal specification of the developing human prefrontal cortex is characterized by the presence of transient belt of MAP2-positive pyramidal cells in the cortical plate*. Federation of European Neuroscience Societies, Forum, Geneva (Abstract).

Judaš, M., & Cepanec, M. (2007). Adult structure and development of the human fronto-opercular cerebral cortex (Broca's region). *Clin. Linguistics Phonetics*, *21*, 975–989.

Judaš, M., Radoš, M., Jovanov-Milošević, N., Hrabač, P., Štern-Padovan, R., & Kostović, I. (2005). Structural, immunocytochemical, and MR imaging properties of periventricular crossroads of growing cortical pathways in preterm infants. *AJNR Am. J. Neuroradiol.*, *26*, 2671–2684.

Kagan, J., & Baird, A. (2004). Brain and behavioral development during childhood. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 93–103). Cambridge, MA: MIT Press.

Khazipov, R., & Luhmann, H. J. (2006). Early patterns of electrical activity in the developing cerebral cortex of humans and rodents. *Trends Neurosci.*, *29*, 414–418.

Kostović, I. (1984). Development of cytoarchitectonic compartments in the putamen of the human fetal brain. *Verh. Anat. Ges.*, *78*, 301–302.

Kostović, I. (1986). Prenatal development of nucleus basalis complex and related fiber systems in man: A histochemical study. *Neuroscience*, *17*, 1047–1077.

Kostović, I. (1990). Structural and histochemical reorganization of the human prefrontal cortex during perinatal and postnatal life. *Prog. Brain Res.*, *85*, 223–240.

Kostović, I., & Goldman-Rakic, P. S. (1983). Transient cholinesterase staining in the mediodorsal nucleus of the thalamus and its connections in the developing human and monkey brain. *J. Comp. Neurol.*, *219*, 431–447.

Kostović, I., & Jovanov-Milošević, N. (2006). The development of cerebral connections during the first 20–45 weeks' gestation. *Sem. Fetal Neonatal Med.*, *11*, 415–422.

Kostović, I., Jovanov-Milošević, N., Krsnik, Ž., Petanjek, Z., & Judaš, M. (2004/2005). Laminar organization of the marginal zone in the human fetal cortex. *Neuroembryology*, *3*, 19–26.

Kostović, I., & Judaš, M. (2002). Correlation between the sequential ingrowth of afferents and transient patterns of cortical lamination in preterm infants. *Anat. Rec.*, *267*, 1–6.

Kostović, I., & Judaš, M. (2006). Prolonged coexistence of transient and permanent circuitry elements in the developing cerebral cortex of fetuses and preterm infants. *Dev. Med. Child Neurol.*, *48*, 388–393.

Kostović, I., & Judaš, M. (2007). Transient patterns of cortical lamination during prenatal life: Do they have implications for treatment? *Neurosci. Biobehav. Rev.*, *31*, 1157–1168.

Kostović, I., Judaš, M., & Petanjek, Z. (2008). Structural development of the human prefrontal cortex. In C. A. Nelson and M. Luciana (Eds.), *Handbook of Developmental Cognitive Neuroscience*. (pp. 213–235). Cambridge, MA: MIT Press, A Bradford Book.

Kostović, I., Judaš, M., Petanjek, Z., & Šimić, G. (1995). Ontogenesis of goal-directed behavior: Anatomo-functional considerations. *Int. J. Psychophysiol.*, *19*, 85–102.

Kostović, I., Judaš, M., Radoš, M., & Hrabač, P. (2002). Laminar organization of the human fetal cerebrum revealed by histochemical markers and magnetic resonance imaging. *Cereb. Cortex*, *12*, 536–544.

Kostović, I., Judaš, M., Škrablin-Kučić, S., Štern-Padovan, R., & Radoš, M. (2006). In vivo MR imaging of transient subplate zone in the human fetal telencephalon. *Soc. Neurosci. Abstracts*, *96*, 10.

Kostović, I., & Krmpotić, J. (1976). Early prenatal ontogenesis of the neuronal connections in the interhemispheric cortex of the human gyrus cinguli. *Verh. Anat. Ges.*, *70*, 305–316.

Kostović, I., Lukinović, N., Judaš, M., Bogdanović, N., Mrzljak, L., Zečević, N., et al. (1989). Structural basis of the developmental plasticity in the human cerebral cortex: The role of the transient subplate zone. *Metab. Brain Dis.*, *4*, 17–23.

Kostović, I., & Molliver, M. E. (1974). A new interpretation of the laminar development of cerebral cortex: Synaptogenesis in different layers of neopallium in the human fetus. *Anat. Rec.*, *178*, 395.

Kostović, I., Petanjek, Z., Delalle, I., & Judaš, M. (1992). Developmental reorganization of the human association cortex during the perinatal and postnatal life. In I. Kostović, S. Knežević, G. Spillich, & J. Wisniewski (Eds.), *Neurodevelopment, Aging, and Cognition* (pp. 3–17). Boston: Birkhäuser.

Kostović, I., & Rakic, P. (1984). Development of prestriate visual projections in the monkey and human fetal cerebrum revealed by transient cholinesterase staining. *J. Neurosci.*, *4*, 25–42.

Kostović, I., & Rakic, P. (1990). Developmental history of the transient subplate zone in the visual and somatosensory cortex of the macaque monkey and human brain. *J. Comp. Neurol.*, *297*, 441–470.

Kostović, I., Škavić, J., & Strinović, D. (1988). Acetylcholinesterase in the human frontal associative cortex during the period of cognitive development: Early laminar shifts and late innervation of pyramidal neurons. *Neurosci. Lett.*, *90*, 107–112.

Kostović, I., Štfulj-Fučić, A., Mrzljak, L., Jukić, S., & Delalle, I. (1991). Prenatal and perinatal development of the somatostatin-immunoreactive neurons in the human prefrontal cortex. *Neurosci. Lett.*, *124*, 153–156.

Krmpotić-Nemanić, J., Kostović, I., Vidić, Z., Nemanić, D., & Kostović-Knežević, L. J. (1987). Development of Cajal-Retzius cells in the human auditory cortex. *Acta Otolaryngol. (Stockh.)*, *103*, 477–480.

La^{m}antia, A. S., & Rakic, P. (1990). Axon overproduction and elimination in the corpus callosum of the developing rhesus monkey. *J. Neurosci.*, *10*, 2156–2175.

Letinić, K., & Kostović, I. (1996). Transient patterns of calbindin-D28k expression in the developing striatum of man. *Neurosci. Lett.*, *220*, 211–214.

Levitt, P. (2003). Structural and functional maturation of the developing primate brain. *J. Pediatr.*, *143*, S35–S45.

Lewis, D. A., Hashimoto, T., & Volk, D. W. (2005). Cortical inhibitory neurons and schizophrenia. *Nat. Rev. Neurosci.*, *6*, 312–324.

Lund, J. S., & Lewis, D. A. (1993). Local circuit neurons of developing and mature macaque prefrontal cortex: Golgi and immunocytochemical characteristics. *J. Comp. Neurol.*, *328*, 282–312.

Marin-Padilla, M. (1983). Structural organization of the human cerebral cortex prior to the appearance of the cortical plate. *Anat. Embryol. (Berl.)*, *168*, 21–40.

Meinecke, D. L., & Rakic, P. (1992). Expression of GABA and GABA-A receptors by neurons of the subplate zone in developing primate occipital cortex: Evidence for transient local circuits. *J. Comp. Neurol.*, *317*, 91–101.

Meyer, G., & Goffinet, A. M. (1998). Prenatal development of reelin-immunoreactive neurons in the human neocortex. *J. Comp. Neurol.*, *397*, 29–40.

Meyer, G., Schaaps, J. P., Moreau, L., & Goffinet, A. M. (2000). Embryonic and early fetal development of the human neocortex. *J. Neurosci.*, *20*, 1858–1868.

Molliver, M. E. (1967). An ontogenetic study of evoked somesthetic cortical responses in the sheep. *Prog. Brain Res.*, *26*, 78–91.

Molliver, M. E., Kostović, I., & van der Loos, H. (1973). The development of synapses in cerebral cortex of the human fetus. *Brain Res.*, *50*, 403–407.

Mrzljak, L., Uylings, H. B. M., Kostović, I., & van Eden, C. G. (1988). Prenatal development of neurons in the human prefrontal cortex. I: A qualitative Golgi study. *J. Comp. Neurol.*, *271*, 355–386.

Mrzljak, L., Uylings, H. B. M., Kostović, I., & van Eden, C. G. (1992). Prenatal development of neurons in the human prefrontal cortex. II: A quantitative Golgi study. *J. Comp. Neurol.*, *316*, 485–496.

Mrzljak, L., Uylings, H. B. M., van Eden, C. G., & Judaš, J. (1990). Neuronal development in human prefrontal cortex in prenatal and postnatal stages. *Prog. Brain Res.*, *85*, 185–222.

Nikolić, I., & Kostović, I. (1986). Development of the lateral amygdaloid nucleus in the human foetus: Transient presence of discrete cytoarchitectonic units. *Anat. Embryol. (Berl.)*, *174*, 355–360.

Nobin, A., & Björklund, A. (1973). Topography of the monoamine neuron systems in the human brain as revealed in fetuses. *Acta Physiol. Scand.*, *388* (Suppl.), 1–40.

Novak, G. P., Kurtzberg, D., Kreuzer, J. A., & Vaughan, H. G., Jr. (1989). Cortical responses to speech sounds and their formants in normal infants: Maturational sequence and spatiotemporal analysis. *Electroencephalogr. Clin. Neurophysiol.*, *73*, 295–305.

O'Leary, D. D. M., Chou, S. J., & Sahara, S. (2007). Area patterning of the mammalian cortex. *Neuron*, *56*, 252–269.

Penn, A. A., & Shatz, C. J. (1999). Brain waves and brain wiring: The role of endogenous and sensory-driven neural activity in development. *Pediatr. Res.*, *45*, 447–458.

Petanjek, Z., Judaš, M., Kostović, I., & Uylings, H. B. M. (2008). Lifespan alterations of basal dendritic trees of pyramidal neurons in the human prefrontal cortex: A layer-specific pattern. *Cereb. Cortex*, *18*, 915–929.

Petrides, M., & Pandya, D. N. (2007). Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *J. Neurosci.*, *27*, 11573–11586.

Preuss, T. M. (2004). What is it like to be a human? In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences*, 3rd ed. (pp. 5–22). Cambridge, MA: MIT Press, A Bradford Book.

Rakic, P. (1977). Prenatal development of the visual system in the rhesus monkey. *Philos. Trans. R. Soc. London [Biol.]*, *278*, 245–260.

Rakic, P. (1988). Specification of cerebral cortical areas. *Science*, *241*, 170–176.

Rakic, P. (2006). A century of progress in corticoneurogenesis: From silver impregnation to genetic engineering. *Cereb. Cortex*, *16*, i3–i17.

Rakic, P., & Zecevic, N. (2003). Emerging complexity of layer I in human cerebral cortex. *Cereb. Cortex*, *13*, 1072–1083.

Schwartz, M. L., Rakic, P., & Goldman-Rakic, P. S. (1991). Early phenotype expression of cortical neurons: Evidence that a subclass of migrating neurons have callosal axons. *Proc. Natl. Acad. Sci. USA*, *88*, 1354–1358.

Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal mapping of cortical thickness and brain growth in normal children. *J. Neurosci.*, *24*, 8223–8231.

Tolonen, M., Palva, J. M., Andersson, S., & Vanhatalo, S. (2007). Development of the spontaneous activity transients and ongoing cortical activity in human preterm babies. *Neuroscience*, *145*, 997–1006.

Vanhatalo, S., & Kaila, K. (2006). Development of neonatal EEG activity: From phenomenology to physiology. *Sem. Fetal Neonatal Med.*, *11*, 471–478.

Vanhatalo, S., Palva, J. M., Andersson, S., Rivera, C., Voipio, J., & Kaila, K. (2005). Slow endogenous activity transients and developmental expression of K^+-Cl^- cotransporter 2 in the immature human cortex. *Eur. J. Neurosci.*, *22*, 2799–2804.

Verney, C., Lebrand, C., & Gaspar, P. (2002). Changing distribution of monoaminergic markers in the developing human cerebral cortex with special emphasis on the serotonin transporter. *Anat. Rec.*, *267*, 87–93.

Voigt, T., Opitz, T., & DeLima, A. D. (2001). Synchronous oscillatory activity in immature cortical network is driven by GABAergic preplate neurons. *J. Neurosci.*, *21*, 8895–8905.

von Economo, C., & Koskinas, G. N. (1925). *Die Cytoarchitektonik der Hirnrinde des erwachsenen Menschen*. Vienna: Verlag von Julius Springer.

von Monakow, C. (1905). *Gehirnpathologie*. Vienna: Alfred Hölder.

Vukšić, M., Radoš, M., & Kostović, I. (2008). Structural basis of developmental plasticity in the corticostriatal system. *Collegium Antropologicum*, *32* (Suppl. 1), 155–159.

Werker, J. F., & Vouloumanos, A. (2001). Speech and language processing in infancy: A neurocognitive approach. In C. A. Nelson and M. Luciana (Eds.), *Handbook of Developmental Cognitive Neuroscience* (pp. 269–280). Cambridge, MA: MIT Press, A Bradford Book.

Yakovlev, P. I., & Lecours, A. R. (1967). The myelogenetic cycles of regional maturation of the brain. In A. Minkowski (Ed.), *Regional Development of the Brain in Early Life* (pp. 3–69). Philadelphia: Davis.

Zecevic, N., & Milosevic, A. (1997). Initial development of gamma-aminobutyric acid immunoreactivity in the human cerebral cortex. *J. Comp. Neurol.*, *380*, 495–506.

Zecevic, N., & Verney, C. (1995). Development of the catecholamine neurons in human embryos and fetuses, with special emphasis on the innervation of the cerebral cortex. *J. Comp. Neurol.*, *351*, 509–535.

3 The Cognitive Neuroscience of Human Uniqueness

TODD M. PREUSS

ABSTRACT Until recently, neuroscientists have lacked powerful means for studying the human brain, and so have relied on studies of nonhuman species for understanding human brain organization. Moreover, the Darwin-Huxley claim that the human mind and brain, while highly developed, are qualitatively similar to those of other species encouraged the concentration of research in a very few "model" nonhuman species. Several recent developments challenge the traditional model-animal research paradigm and provide the foundations of a new neuroscience. First, evolutionary biologists now understand that living species cannot be arrayed along a single, unbroken sequence of phylogenetic development: species can differ qualitatively. Second, neuroscientists are documenting remarkable variations in the organization of cerebral cortex and other brain regions across mammals. Third, new noninvasive methods from histology, neuroimaging, and genomics are making the human brain accessible for direct, detailed study as never before. Finally, these same methods are being used to directly compare humans to other species (including chimpanzees, the species most closely related to humans), providing the foundations of a new and richly detailed account of how the human brain both resembles and differs from that of other species.

Evolution isn't what it used to be

What distinguishes the human brain from that of other animals? This is one of the most profound questions that neuroscience confronts, yet surprisingly it has not attracted a great deal of empirical investigation, nor have the answers offered heretofore been particularly revealing. Humans have big brains relative to body size—that much is agreed upon—but what is different about the contents of those big brains?

We should perhaps start by asking why, despite the great advances made in neurosciences over the past few decades, we have so little solid information about how the human brain differs from and resembles that of other animals. One reason, clearly, is that neuroscientists have had relatively limited means for studying humans, compared to the means available for studying nonhuman species. The most powerful methods at our disposal for studying brains—genetic manipulations, tracer injections, microelectrode recording and stimulation—require the use of invasive, and often terminal, experimental procedures. Given this fact, neuroscientific research has tended to focus on nonhuman species. To be sure, neuroscience has maintained a tradition of human research, beginning with clinical neurology and strengthened today by the availability of a suite of remarkable noninvasive imaging tools. Nonetheless, there remains an important gap between human and nonhuman studies: the techniques available for studying animals permit much more detailed investigations involving finer levels of organization, and the use of model animals facilitates better experimental design. The result is that many of our ideas about human brain organization are actually inferences drawn from studies of nonhuman species.

What is the scientific basis for making inferences about human brain organization from the study of nonhuman species? The answer, seemingly, is the principle of evolution, which asserts that there was continuity between the human species and other animals through Earth history. We rightly honor Charles Darwin for having the profound insight that all animals are descended from one (or a few) progenitor species and providing the evidence necessary to convince scientists of the truth of that insight (Darwin, 1859). But Darwin did more than that. Darwin viewed the forces of evolution—mainly natural selection—as means of improvement, so that those species that have been subjected to more selection are better species. In this way, Darwin provided a naturalistic explanation of the Great Chain of Being (*scala naturae*), the idea that life forms can be arranged along a linear scale from the simple and base to the complex and refined. The concept of a scale of being long predated Darwin, having its origins (in the Western tradition, at least) in Aristotelian philosophy (Lovejoy, 1964). (Some pre-Darwinian versions of the scale of being go beyond the merely human, extending through the grades of angels up to the Almighty.) Darwin took the metaphysical Great Chain of Being, which seems very strange to most of us today, and turned it into something familiar—the phylogenetic scale (Richards, 1987, 1992). Accordingly, depictions of evolutionary history from the 1860s through the 1960s commonly represented evolution as a process of ascent, with Man at the top (figures 3.1*A*, 3.2*A*, 3.2*B*). Although some historians have

TODD M. PREUSS Division of Neuroscience and Center for Behavioral Neuroscience, Yerkes Primate Research Center, Emory University, Atlanta, Georgia

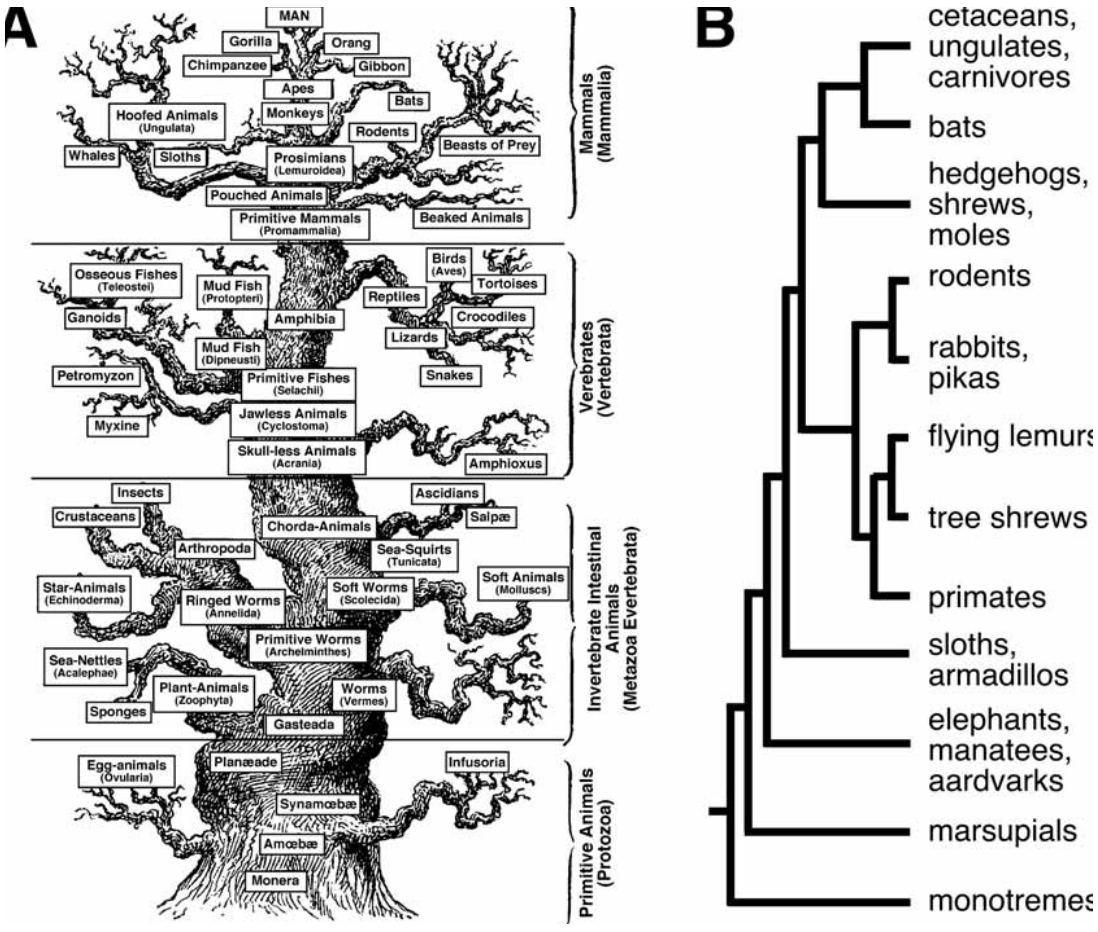


FIGURE 3.1 Changing views of evolutionary history, contrasting (*A*) an early phylogenetic tree from Ernst Haeckel (1874, 1879), an early supporter of Darwin, and (*B*) a modern phylogeny of mammals from recent comparative molecular studies, based on Murphy and colleagues (2001). In Haeckel's tree (*A*), the course of evolution is upward, and it culminates in humans. Other species are relegated to side branches. The modern tree (*B*) has no orientation, except with respect to time, and one locates primates or humans merely from the shape of the tree. The tree is drawn with the present-day time horizontal to the right, but the direction of the time axis is arbitrary. Note that I have translated some of the German terms for various primate groups differently than in the original English edition of Haeckel's work (Haeckel, 1879) to better conform to modern usage.

credited Darwin with the idea that evolution has no direction or orientation, it is clear from even a casual reading of Darwin (see especially Darwin, 1871) that he firmly believed in the phylogenetic scale, and that human beings are at the top—that we are the most advanced form of animal life.

Darwin's work marks the beginning of evolutionary biology, but like other branches of science, evolutionary biology underwent revolutionary changes during the 20th century, resulting in a worldview that Darwin would probably have found at once very familiar and very strange. Detailed studies of evolutionary mechanisms made it clear that adaptation is a local process, fitting populations to their particular circumstances. Adaptation does not necessarily involve increased complexity; in fact, it often results in simplification or loss of structures and functional capacities, as in the case of many parasitic organisms. Yet parasites are arguably just as well adapted to their circumstances as the organisms they infest. Furthermore, evolutionary history has proven to be anything but an unblemished chronicle of progressive improvement: cosmic collisions and tectonic upheavals repeatedly shuffled the deck.

There is another important transformation in the way evolutionary biologists came to view the history of life, and like the study of adaptation, this transformation was ultimately rooted in another of Darwin's profound insights. The sole figure in the *Origin of Species* (Darwin, 1859) depicts evolution as a branching process, with ancestral lineages splitting to form multiple daughter species. Early evolutionists did not recognize any contradiction between the phylogenetic scale and the phylogenetic tree; as a result, early trees typically had a narrow, vertical orientation (figures 3.1*A*; 3.2*A*, 3.2*B*). The contradictions became clear, however, when the lack of a unitary direction to evolutionary history was appreciated. If every lineage is shaped by selection

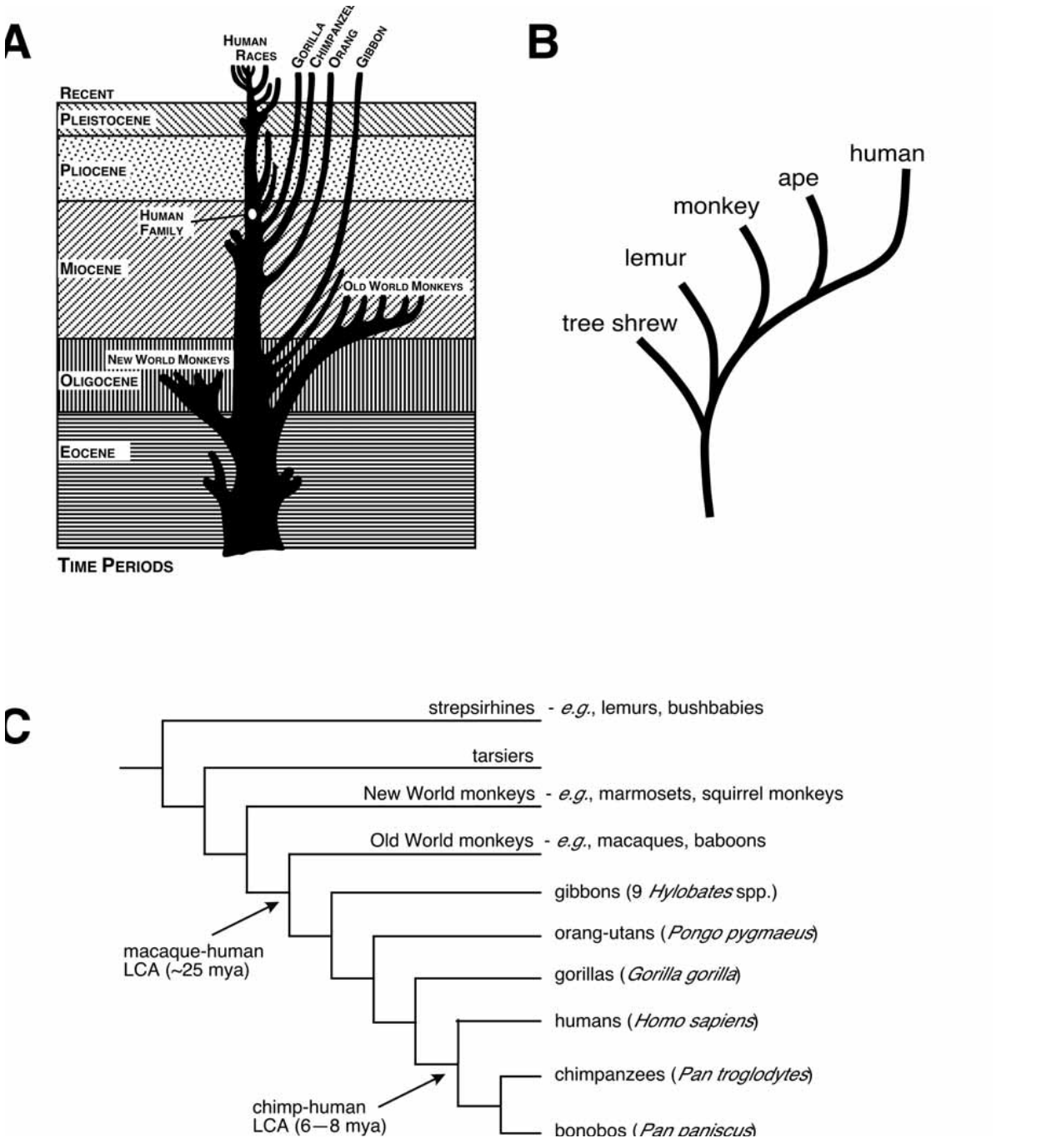


FIGURE 3.2 Changing views of primate evolution, contrasting the vertical, human-oriented trees of (*A*) Grafton Elliot Smith (1924) and (*B*) W. E. Le Gros Clark (1959), with (*C*) a modern branching diagram of the primate order, based primarily on Fleagle (1999). Elliot Smith and Le Gros Clark were both neuroanatomists and primatologists, and their views were enormously influential in shaping our understanding of primate evolutionary history and primate brain evolution (Preuss, 1993, 2007a, 2007b). In their trees, like Haeckel's tree (figure 3.1*A*), evolution is a progression toward a human apex. In Elliot Smith (*A*), even living members of the primate order, like New World and Old World monkeys, are treated as though they were extinct side branches. Le Gros Clark (*B*) considered living forms like tree shrews, lemurs, monkeys, and apes to be the living embodiments of grades of organization or stages of evolution leading to humans. Modern trees (such as *C*; modified from Preuss, Cáceres, Oldham, & Geschwind, 2004) place all living forms on the present-day time horizon, and again, the direction of the time axis is arbitrary. Furthermore, there are many equivalent ways of arraying modern groups along the current time horizon, since tree branches can be rotated around any of the internal nodes. Humans, therefore, can be placed at the top of the list, the bottom of the list, or somewhere in between (as shown in *C*). LCA, last common ancestor.

pressures specific to its circumstances, then the main axis of the tree can be understood to represent time, rather than also serving as an index of advancement. With this recognition, evolutionary trees have become much wider and more rectangular, emphasizing that evolutionary history is a matter of diversification through time (figures 3.1*B*, 3.2*C*).

There has been at least one more important change in evolutionists' view of life. Darwin emphasized the continuity of life, and asserted that discontinuities between humans and our closest living relatives would invalidate the theory of evolution. Accordingly, he proposed a research program that encouraged scientists to seek in nonhuman species simplified rudiments of characteristics that are well developed in humans (Darwin, 1871). Huxley (1863) took a similar approach to the brain. For modern evolutionists, by contrast, continuity is seen in strictly historical terms: species are understood to be connected by the continuity of generations through time. Any set of *living* species can, however, differ in ways that encompass no living intermediates. We know, for example, that cetaceans (whales and dolphins) were thoroughly transformed by evolution from terrestrial quadrupeds (related to modern sheep and cows) to fishlike aquatic forms through a series of intermediate forms, replacing, step by step, anatomical features suited for terrestrial life with features suited to aquatic life. We know this because there happens to be a spectacular fossil record documenting the extinct intermediate forms (Gingerich, Haq, Zalmout, Khan, & Malkani, 2001; Thewissen, Williams, Roe, & Hussain, 2001). But there are no *modern* intermediate forms, no array of living species that spans the gap between the aquatic cetaceans and their terrestrial kin. Thus there is no reason to maintain the narrow interpretation of continuity promulgated by Darwin and Huxley, and no reason to insist that every aspect of the human brain or mind be present in some lesser form in other species (see also Penn, Holyoak, & Povinelli, 2008). In the light of modern evidence about evolutionary change, we can see that Darwin and Huxley confounded two different kinds of continuity.

One of the most important aspects of the modern understanding of evolution, therefore, is the change in how biologists view the similarities and differences among animals. If you are troubled by the term "human uniqueness" in the title of this essay, you're in good company—Darwin, with his emphasis on continuity, would have been troubled, too. But modern evolutionary biology regards each species as a unique outcome of evolutionary history. To be sure, any two species will share some features in common, by virtue of being descended from a common ancestor. But as one restricts comparisons to closer and closer relatives, the set of shared features will become more limited. Humans have forward-facing eyes and nails, rather than claws, on our fingertips; we share these features with other primates, but not with other mammals (unless, like cats, with their forward-facing eyes, they evolved them independently). Finally, every species will have features that evolved in its own lineage after that lineage branched off from the line leading to its closest relatives. These are its species-specific features. Thus every species can be regarded as a mosaic of characteristics, some that are shared with other species and some that are specific to it.

Model-animal versus comparative approaches to neuroscience

When it comes to the issue of how the brains of different species resemble and differ from one another, there is no question that the majority of neuroscientists would stand with Darwin. Most neuroscientists study nonhuman species, and, moreover, most research effort is concentrated on one of the standard model animals, such as *Drosophila*, *C. elegans*, rats, mice, or macaque monkeys. The emphasis on studies of model animals, which comes to us from the biomedical research tradition and from experimental psychology, is predicated on the assumption that there are major or basic features of organization that are widely shared across animals, and that the differences, whatever they are, are relatively minor and unimportant (Logan, 2001, 2005; Preuss, 2000a). Of course, propagandists for different model species argue endlessly about why their model is the best—highlighting the fact that species do, after all, differ—but the assumption that we can get at the basics by studying a few species is rarely questioned. Neuroscientists working in the zoological tradition tend to take a broader, more comparative view (e.g., Striedter, 2005), but that approach has been marginalized in the neurosciences as the model-animal paradigm has become the accepted approach to biomedicine. Most neuroscientists simply take generality for granted, in effect treating it as a property of the models themselves. Modern publications reporting results in rats or mice typically do not even identify the species in the title. Common model animals like fruit flies, rats, and mice are no longer treated as subjects so much as standardized reagents (Logan, 2002, 2005).

The assumption of generality, and the resulting concentration of research effort in a few model species, seemed until recently to rest on a solid empirical foundation. Brain organization was regarded as evolutionarily conservative, at least across mammals. Mammalian brains might vary in size, and possibly also in degree of "differentiation" (a term meaning quite different things to different people), but the internal organization of the brain, especially at the level of the local cellular architecture and local connectivity, was understood to be largely invariant. Cortex, in particular, was said to exhibit a "basic uniformity" across species (Rockel, Hiorns, & Powell, 1980; see also Creutzfeldt, 1977; Mountcastle, 1978; Phillips, Zeki, & Barlow, 1983; Szentagothai, 1975).

During the 1970s, however, evidence began to accumulate indicating that at least the larger-scale organization of cerebral cortex was not so uniform across species: different mammalian groups evolved different complements of cortical areas (e.g., Allman, 1977; Kaas, 1977; see also Allman, 1999; Kaas, 2007). As comparative studies of connectivity have advanced, it has become evident that the way cortical areas are organized into functional networks can vary markedly across species (reviewed by Preuss, 2007a, 2007b). In recent years, too, the empirical basis for the claim of basic uniformity at local levels of organization has collapsed, with variations in cell phenotypes, local connectivity, laminar organization, and modular organization proving to be ubiquitous. Even among closely related mammalian species, there can be marked differences at these finer levels of cortical organization (for reviews, see DeFelipe et al., 2007; DeFelipe, Alonso-Nanclares, & Arellano, 2002; Elston, 2007; Hof & Sherwood, 2005, 2007; Preuss, 1995, 2001; Sherwood & Hof, 2007).

The implications of these results for the practice of neuroscience and for our understanding of the human brain are profound. Since brains are diverse, any attempt to discover general or shared features of organization requires comparative studies—generalizing from just a few "standard" species just will not do. If we can acknowledge the pitfalls of extrapolating results from one mouse strain to another, or from mice to rats (Cantallops & Routtenberg, 2000; McNamara, Namgung, & Routtenberg, 1996; Rekart, Sandoval, & Routtenberg, 2007), should we not be at least as mindful of the pitfalls of extrapolating from mice to humans? Evolutionary biology provides standards and methods for rigorous comparative studies, methods that have been adopted in many areas of biology—comparative genomics, for example, which would be impossible without the analytic tools of modern phylogenetics—but these methods are still unknown to most neuroscientists.

No less problematic than the concentration of so much research in so few species is the limited research directly comparing humans to other species. If one accepts the modern view of evolution, then it is clear there must be ways that humans are like other animals and ways that we are distinctive. How can we identify and distinguish these similarities and differences except by means of comparative studies of humans and other animals? Furthermore, if we want to understand in detail what is distinctively human about our brains, the most informative approach is to compare humans to the animals that are most closely related to us: the other hominoids. The primate superfamily Hominoidea consists of humans, chimpanzee, bonobos, gorillas, orangutans, and gibbons: Collectively, this group of lineages (or "clade") is the sister group of Old World monkeys (the clade that includes macaque monkeys), the two groups sharing a common ancestor about 25 million years ago (figure 3.2*C*). The lineage most closely related to humans is the chimpanzee-bonobo clade, with which we shared a common ancestor about 6–8 million years ago. If we take evolutionary biology seriously, we should expect that hominoids, as a group, possess brain characteristics that are absent in Old World monkeys (and vice versa), and that each hominoid species—humans included—should possess brain features lacking in the other species.

Human brain specializations: New methods, new discoveries

What, then, do we know about how human brains differ from those of chimpanzees and other hominoids? Until recently, one could answer this question quite succinctly: not much. One fact that is undisputed is that humans have very large brains, even when body size is factored out: humans are only slightly larger than chimpanzees in average adult body size, but human brains average around 1,200–1,300 cubic centimeters (cc), compared to a little less than 400 cc for chimpanzees (de Sousa & Wood, 2007; Kaas & Preuss, 2008). Most of this increase occurred during the last 2 million years, after the genus *Homo* evolved from *Australopithecus*-like ancestors, embarking on a lifestyle that involved reliance on stone tools and a diet that included an important component of animal flesh.

Beyond these facts, there have been few points of general agreement. Since certain human brain functions appear to be highly lateralized compared to nonhuman primates (Corballis, 2007), it's natural to suppose that the well-documented anatomical asymmetries of the human brain, such as the larger extent of the planum temporale in the left hemisphere than in the right (Geschwind & Levitsky, 1968), must also be human-specific. Apes, however, possess at least some of the asymmetries seen in humans, including the asymmetry of the planum temporale (e.g., Gannon, Holloway, Broadfield, & Braun, 1998; Hopkins, Marino, Rilling, & MacGregor, 1998). This is a surprising result, since the planum is involved in language and is sometimes identified with Wernicke's area. Do humans, with our greatly enlarged brains, possess cortical areas in addition to those present in apes and other nonhuman primates? It seems plausible that they do—certainly Brodmann (1909) and others (e.g., Crick & Jones, 1993) have supposed that humans must possess new areas to support new, human-specific brain functions. Modern cortical mapping studies, however, suggest mainly commonalities in the complement of cortical areas in humans and macaques (Petrides & Pandya, 1994), the latter being the most widely studied nonhuman primates. There is evidence, furthermore, that nonhuman primates possess homologues of Broca's and Wernicke's areas (Preuss, 2000b), which suggests that the evolution of new human functions did not necessarily require the addition of new cortical areas. In settling

the general question of new areas, however, we remain hamstrung by the lack of modern, comprehensive brain maps for humans and other hominoids—the state-of-the-art chimpanzee cortical map is nearly 60 years old (P. Bailey, Bonin, & McCulloch, 1950). At present, there is not even unanimity about whether the evolutionary expansion of the human brain involved the disproportionate expansion of the association cortex or whether expansion was more nearly global (see the review of Rilling, 2006), although I feel the evidence strongly supports the former view (Preuss, 2004b). Other, no less fundamental, aspects of human brain organization have received even less attention. In some cases, like the local cellular and histological organization of the brain, the lack of attention probably reflects simple indifference: if cortical organization is uniform across species, as has been supposed, why waste one's time looking for differences? In other cases, the lack of attention probably reflects the lack of appropriate techniques: if studying brain connectivity requires invasive and terminal procedures, then you cannot study connectivity in humans, nor can you do so in other hominoids, which are also off-limits for invasive neuroscience research.

Fortunately, the development of new methodologies, combined with new ways of thinking, has begun to transform our understanding of the human brain. For the sake of discussion, I will group the methodologies into three categories—comparative histology, comparative neuroimaging, and comparative genomics. Although seemingly a very disparate group of techniques, they have in common the quality that they do not require invasive and terminal experimental procedures, and for this reason they can all be used to study humans directly and to compare humans to chimpanzees or to just about any other species. It is their noninvasive character that makes these techniques so extraordinarily valuable for understanding human brain specializations.

Comparative Histology One does not normally associate the word "histology" with "new," but in fact there is much that is new in histology, and it has contributed some of the most remarkable recent discoveries about hominoid and human brains. Histology today has at its disposal a fantastic set of tools for localizing specific molecules within the brain—antibodies, lectins, riboprobes, as well as conventional histochemical stains—and because specific molecules tend to be associated with specific cell compartments, cell types, laminae, modules, and areas, these tools make possible a molecular dissection of brain structure. In addition, there are new stereological methods for counting cells and other elements in the brain that constitute a major advance over traditional counting methods (Schmitz & Hof, 2005; West, 1999). Finally, there are new storage solutions that are far superior to formalin for maintaining the long-term viability of tissue for molecular analysis (Hoffman & Le, 2004). Modern cryopreservative solution greatly enhances the value of tissue archives, so studies can make use of tissue harvested from individuals who die of natural causes, and need not rely entirely on tissue obtained from terminal experiments.

What we have learned using these techniques, in the context of comparative studies, is that hominoids and humans possess specialized neuronal phenotypes and specialized columnar, modular, and laminar arrangements of nerve cells (Hof & Sherwood, 2007; Preuss, 2004a; Sherwood & Hof, 2007). A particularly dramatic example is the discovery of a distinctive class of large, spindle-shaped neurons in layer 5 of anterior cingulate and frontoinsular cortex, termed "spindle cells" (Nimchinsky et al., 1999) or "von Economo neurons" (Allman, Watson, Tetreault, & Hakeem, 2005). These are present in humans and other hominoids (except gibbons), but are much larger and more numerous in humans. They are absent in other primate and most other mammalian species examined to date. In a clear case of convergent evolution, however, large spindle-shaped cells are present in homologous regions of some large-brained cetaceans (Hof & Van Der Gucht, 2007). It has been proposed that spindle cells evolved to facilitate rapid transmission of emotion- and reward-relevant information in large-brained, highly social hominoids and cetaceans (Allman et al.; Watson & Allman, 2007).

Many additional neuronal and neuritic specializations have been described in hominoids. These include an additional distinctive class of layer 5 pyramidal cell in anterior cingulate cortex, specialized by virtue of expressing calretinin, a calcium-binding protein expressed only in interneurons in most mammalian groups (Hof, Nimchinsky, Perl, & Erwin, 2001). Raghanti and colleagues have described hominoid specializations of the density and morphologies of serotonergic (Raghanti et al., 2008b) and cholinergic fibers (Raghanti et al., 2008a) in frontal cortex, including areas 9 and 32.

There have been fewer comparative studies of the microarchitecture of human language cortex than one might have supposed. Although apes and humans apparently have similar morphological asymmetries in Wernicke's area, there appear to be asymmetries at finer levels of organization. Buxhoeveden, Casanova, and their colleagues report that the spacing of cellular columns in cortical area Tpt, located on the planum temporale, is asymmetrical in humans, being wider on the left than the right, but not in chimpanzees and macaques (Buxhoeveden, Switala, Litaker, Roy, & Casanova, 2001).

While we would expect to find changes in brain regions that support functions known to have changed in human evolution, it is significant that the changes are widespread and include regions not conventionally thought to be evolutionarily "advanced." Thus the spindle cells and the morphological specializations of serotonergic and cholinergic fibers noted above are present in limbic cortex (areas 24

and 32). Nor are specializations necessarily limited to cortex: the basal forebrain magnocellular neurons of New World and Old World monkeys express the peptide galanin in the cytoplasm, whereas those of apes and humans do not; moreover, the basal forebrain of hominoids, but not monkeys, exhibits plexus of galanin-containing processes of extrinsic origin (Benzing, Kordower, & Mufson, 1993; Kordower & Mufson, 1990).

Currently, we know about as little about the cognitive and behavioral specializations of humans as we do about human brain specializations (Subiaul, Barth, Okamoto-Barth, & Povinelli, 2007) and thus should be alert to the possibility that specializations will turn up in places we do not expect, including sensory systems. A case in point: In 1999, my laboratory discovered, quite by accident, that the histology of human primary visual cortex (area V1) differs strikingly from that of apes or any monkey species that has been examined (Preuss, Qi, & Kaas, 1999). We documented a distinctive modular organization of human layer 4A, with cell bodies and neurites that label with SMI-32 and MAP2 antibodies (markers for neurofilaments and microtubules, respectively) distributed in a meshlike arrangement, with the mesh surrounding unlabeled territories. Subsequently, we found that the mesh is densely immunoreactive for Cat-301 (figure 3.3), an antibody considered a selective marker for elements of the magnocellular (M) pathway of the visual system, and that the intervening territories contain small cells that strongly express calbindin, a calcium-binding protein (Preuss & Coleman, 2002). We have speculated that the cortical processing of motion and contrast information, which is mediated by the M pathway, was modified in human evolution; additional evidence for such changes comes from studies of higher-order visual areas (discussed later). Whether or not this particular functional interpretation is correct, the structure of layer 4A of humans is very different from that of other primates. No other primate examined to date possesses patches of strong labeling in layer 4A with SMI-32, or MAP2, or Cat-301. Like humans, chimpanzees display dense calbindin expression in layer 4A, but calbindin is distributed homogenously in chimpanzees rather than in the patchy manner characteristic of humans. In Old World and New World monkeys, layer 4A is calbindin poor, although some other layers express it strongly. As it happens, this is just one example of many variations in the organization of the primary visual cortex, variations that distinguish humans from other hominoids, hominoids from macaque monkeys, and different monkey species from each other (Preuss, 2004a).

Comparative Neuroimaging The foregoing discussion illustrates the power of histological techniques to resolve differences in the microarchitecture of the brain between even very closely related species. New neuroimaging techniques make it possible to compare the connectivity and regional organization of the brain, and so provide insight into the evolution of the higher-order organization of the brain. As with the histological techniques, the imaging techniques are noninvasive, and so make it possible to compare humans and nonhuman primates on something like a level playing field.

The earliest applications of comparative imaging involved structural magnetic resonance imaging (MRI) scans (usually T1-weighted scans) to study morphometry—quantifying the sizes and proportions of the different cerebral lobes across species (Rilling & Insel, 1999b; Rilling & Seligman, 2002; Semendeferi & Damasio, 2000; Semendeferi, Damasio, Frank, & Van Hoesen, 1997; Semendeferi, Lu, Schenker, & Damasio, 2002), the size of the cerebellum (MacLeod, Zilles, Schleicher, Rilling, & Gibson, 2003; Marino, Rilling, Lin, & Ridgway, 2000), the sizes of white matter structures (Rilling & Insel, 1999a; Schenker, Desgouttes, & Semendeferi, 2005; Schoenemann, Sheehan, & Glotzer, 2005), and patterns of interhemispheric asymmetries (Gilissen, 2001; Hopkins et al., 1998). In principle, there is little in this morphometric work that could not have been done with standard histological methods: sectioning and staining brains, and then taking measurements on the sections. In practice, however, MRI morphometrics offers many advantages: it can make use of live individuals, and even when fixed brains are scanned, data acquisition and analysis are much faster and less labor intensive than with histological material, and the fixed brains remain intact after examination so they can be put to other uses. The result is that MRI facilitates study of a wider range of species than would be practical were the same questions to be addressed histologically.

As noted earlier, the lack of information about brain connectivity in species such as humans and chimpanzees that cannot be studied with invasive tracing techniques has been one of the greatest obstacles to understanding brain organization in those species. The situation has improved considerably in recent years with the introduction of diffusion-tensor imaging (DTI) (Conturo et al., 1999; Mori & van Zijl, 2002; Mori & Zhang, 2006; Ramnani, Behrens, Penny, & Matthews, 2004). DTI measures the aggregate direction and magnitude of water diffusion in brain voxels. Since water tends to diffuse along, rather than across, the hydrophobic myelin sheaths of axons, the direction of water diffusion reflects the direction of fibers in a voxel. With this information, it is possible to reconstruct fiber tracts through the white matter between regions of interest in the brain. Compared to traditional chemical techniques, DTI has significant limitations: its spatial resolution is too coarse to track from neuron to neuron; it does not work well in gray matter, where fiber coherence is low; and it is vulnerable to false positives and false negatives. While studies designed to evaluate what DTI can and cannot reliably determine are continuing, it is apparent that, at least for certain pathways,

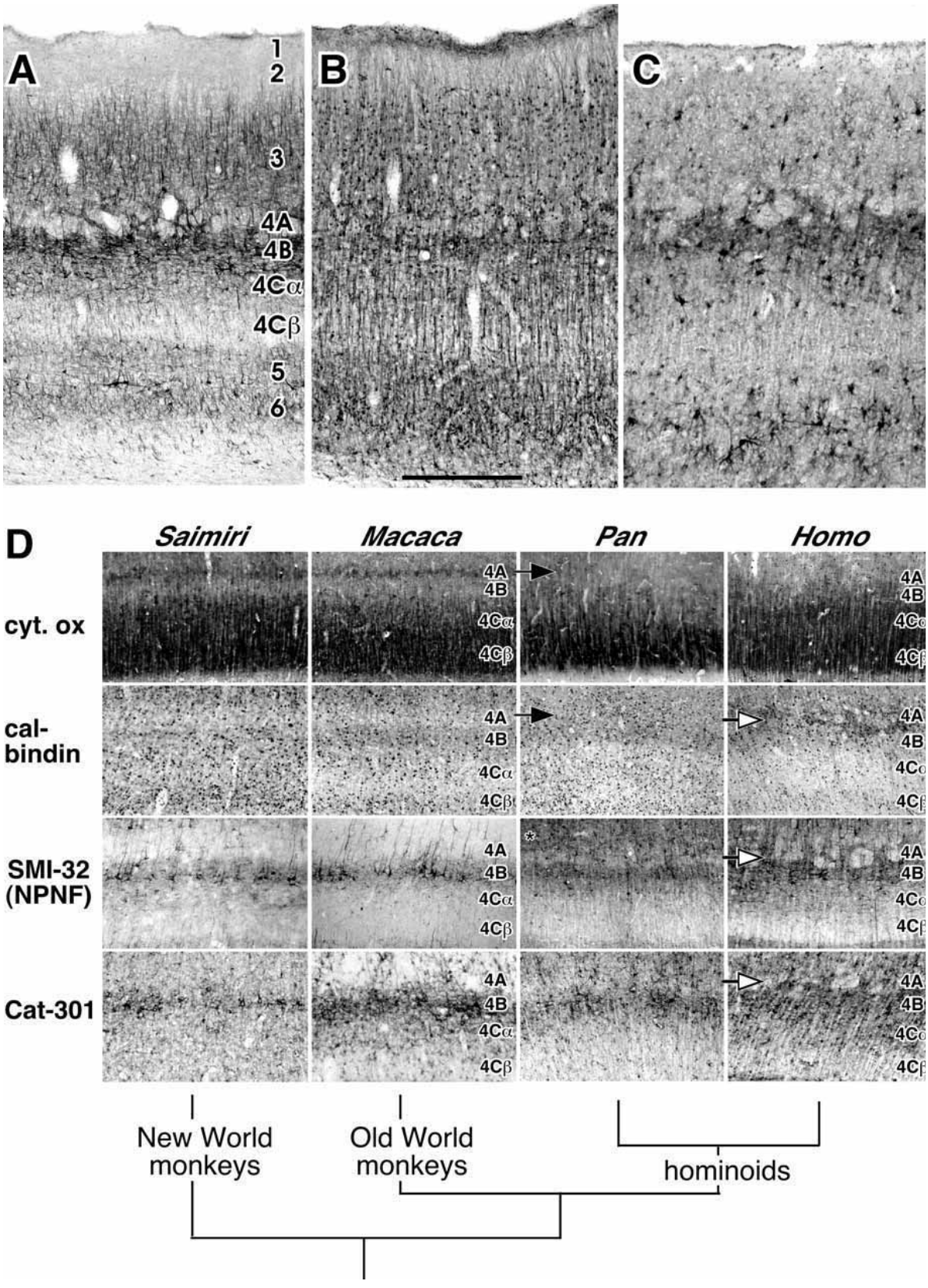

A
1
2
3
4A
4B
4Cα
4Cβ
5
6
B
C
D
Saimiri
Macaca
Pan
Homo
cyt. ox
cal-
bindin
SMI-32
(NPNF)
Cat-301
4A
4B
4Cα
4Cβ
New World
monkeys
Old World
monkeys
hominoids

FIGURE 3.3 Evolutionary modification of human primary visual cortex according to Preuss and Coleman (2002). (*A–C*) Human primary visual cortex labeled for nonphosphorylated neurofilaments with the SMI-32 antibody (*A*), an antibody to MAP2 (*B*), and antibody Cat-301 (*C*), which labels the proteoglycan-rich extracellular matrix. All three preparations show the distinctive compartmental pattern of human layer 4A. Scale = 500 microns. (*D*) Comparative analysis of evolutionary changes in layer 4A and 4B of hominoid and human primary visual cortex. Sections from a New World monkey (*Saimiri*, squirrel monkey), an Old World monkey (*Macaca*), and two hominoids (*Pan*, chimpanzee; *Homo*, human), were labeled with four different preparations: cytochrome oxidase (cyt. ox.), antibodies for calbindin, the SMI-32 antibody for nonphosphorylated neurofilaments, and the Cat-301 antibody for extracellular matrix proteoglycan. The evolutionary relationships between the four taxa are represented by the branching diagram at the bottom of the figure. The ancestors of all four groups probably possessed a cytochrome-oxidase-dense band in layer 4A, a band subsequently lost in hominoid evolution, as it is absent in *Pan* and *Homo*, as indicated by the black arrow. This difference may reflect the reduction or loss of projections to layer 4A from the parvocellular layers of the lateral geniculate nucleus. Calbindin immunoreactivity is weak in layer 4A in *Saimiri* and *Macaca*, but strong in *Pan* and *Homo*, as indicated by the black arrow, suggesting that increased calbindin expression in this layer is a hominoid specialization. (Hominoids also show much stronger staining of the cortex superficial to layer 4A, as indicated by the asterisk in *Pan*.) The distinctive, patchy compartmentation of human layer 4A can be seen in the calbindin, SMI-32, and Cat-301 preparations, as denoted by the white arrows in *Homo*.

accurate reconstructions can be obtained (e.g., Behrens, Berg, Jbabdi, Rushworth, & Woolrich, 2007; Conturo et al., 1999; Dauguet et al., 2007; Parker et al., 2002; Schmahmann et al., 2007).

The advantage of DTI, of course, is that it can be used to study connectivity noninvasively, so it can be used with humans, chimpanzees, and other primates. It can even be used with fixed brains, which makes a wide variety of species accessible for connectivity studies for the first time (e.g., Kaufman, Ahrens, Laidlaw, Zhang, & Allman, 2005). Some of the first comparative studies undertaken with DTI have compared humans and macaque monkeys, examining thalamocortical (Croxson et al., 2005), corticopontine (Ramnani et al., 2006), and prefrontal cortex (Croxson et al.) connectivity. Recently, Rilling and colleagues published the first comparative study of humans, chimpanzees, and macaques, focusing on the arcuate fasciculus (AF), a white matter tract that in humans interconnects Broca's and Wernicke's areas (Geschwind, 1970; Glasser & Rilling, 2008). Rilling and colleagues (2008) tracked pathways between the posterior superior temporal lobe (Wernicke's area) and inferior frontal cortex (Broca's area) in all three species, but only chimpanzees and humans were found to have a distinct AF. Moreover, only in humans did the AF consistently include fibers tracking to the middle temporal gyrus, in addition to the fibers between Broca's and Wernicke's areas (figure 3.4). Functional imaging studies in humans (reviewed by Glasser & Rilling, 2008) indicate that the cortex of the middle temporal gyrus is involved in representing word meaning. Noting evidence that inferotemporal visual cortex appears to be situated more posteriorly and inferiorly in humans than in macaques, Rilling and colleagues suggest that the cortex of the middle temporal gyrus was enlarged and modified in human evolution, and may constitute an evolutionary novelty.

In addition to comparative structural neuroimaging, it is possible to do comparative functional imaging, using functional MRI (fMRI) or positron-emission tomography (PET). Functional MRI is now routine in humans, but the need to prevent head movement in the scanner has meant that its application in nonhuman primates has largely been restricted to macaque monkeys, which can be physically restrained. Although many macaque studies have employed these animals in the role of human models, some workers have made a point of documenting differences between humans and macaques. These studies have identified macaque-human differences in the responsiveness of contrast- and motion-related regions of the dorsal extrastriate visual cortex and posterior parietal cortex (Denys et al., 2004; Orban, Claeys, et al., 2006; Orban et al., 2003; Orban, Van Essen, & Vanduffel, 2004; Tootell et al., 1997; Vanduffel et al., 2002), the differences being sufficient to complicate the interpretation of human-macaque homologies (Sereno & Tootell, 2005) and to suggest that humans possess areas macaques lack (Orban, Claeys, et al.).

It is unlikely that we will soon have fMRI studies of awake, behaving chimpanzees, as imaging equipment is no match for these enormously powerful animals. Yet it is possible to do functional scanning in chimpanzees, using a modification of the 18F-fluroodeoxyglucose (FDG)-PET technique. In this approach, originally developed for macaques, the experimenter provides the animal with the FDG at the start of testing session, and then, after the level of FDG plateaus (45–60 minutes), the animal is anesthetized and PET scanned. This technique was recently used to compare awake, resting-state brain activity in humans and chimpanzees (Rilling et al., 2007). The results demonstrated commonalities between species: both exhibited activation in medial frontal and posteromedial cortices, components of the "default-mode" network thought to be involved in emotion-laden recollection and mental self-projection (e.g., Buckner & Carroll, 2007; Buckner & Vincent, 2007; Gusnard & Raichle, 2001; Raichle & Snyder, 2007). Humans, but not chimpanzees, however, also showed activation of lateral cortical regions in the left hemisphere associated with language and conceptual representation.

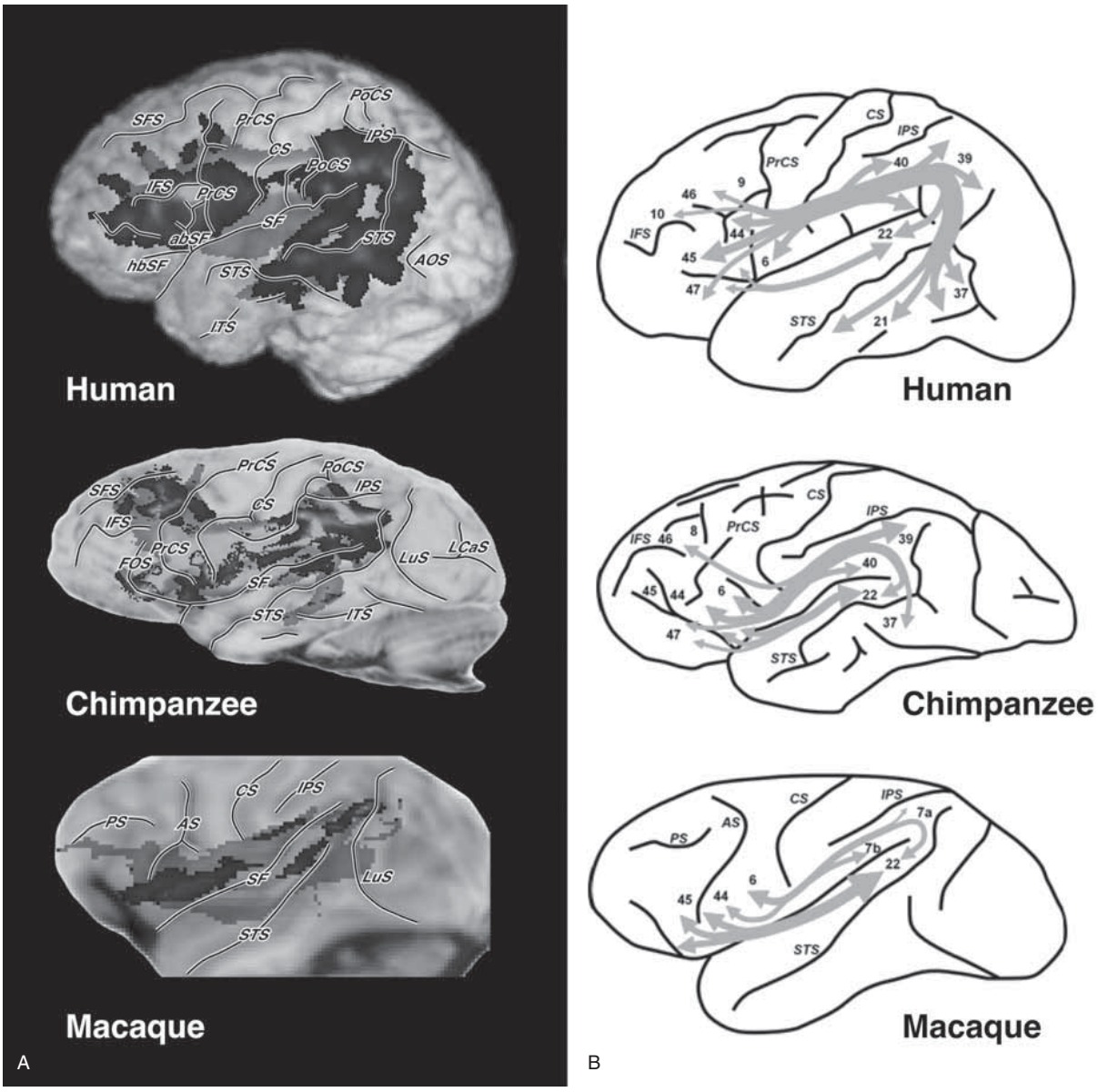


FIGURE 3.4 Evolution of the human arcuate fasciculus (AF), which interconnects frontal and temporal language areas, based on the comparative diffusion-tensor imaging (DTI) results of Rilling and colleagues (2008). (*A*) Average tractography results from the left hemispheres from 10 humans, three chimpanzees, and two macaque monkeys. (*B*) Schematic representation of the results shown in *A*, representing the cortical endpoints of the tracts in terms of Brodmann's areas. Both humans and chimpanzees have a distinct AF, although in humans the AF includes strong connections with the temporal cortex below the superior temporal sulcus (STS), including area 21, a region where word meaning is represented. Chimpanzees have very few fibers in the AF that extend to the cortex inferior to the STS. Macaques do not have a definite AF: fibers traveling between the posterior inferior frontal cortex and posterior temporal lobe take a more ventral route, passing deep to the insula, and include few if any fibers with endpoints inferior to the STS. (See color plate 1.)

Although fMRI scanning of awake chimpanzees currently seems impractical, it may be valuable do fMRI scanning of anesthetized chimpanzees. Although cortical activity is reduced with light anesthesia, it is not eliminated. Patterns of regional coactivation that reflect patterns of anatomical connectivity can be ascertained using so-called functional connectivity MRI (fcMRI), even under light anesthesia (Vincent et al., 2007), and possibly the same can be done with chimpanzees. This technique would provide a valuable addition to DTI as a source of information for exploring the evolution of human brain networks.

COMPARATIVE GENOMICS In recent years, few branches of science have so firmly captured the public imagination—

and that of scientists, too—as the field of comparative genomics, particularly as it pertains to the genetic differences that distinguish human beings from chimpanzees. The story line is compelling: humans and chimpanzees are nearly identical genetically—98–99% similar, apparently (King & Wilson, 1975; Marks, 2002)—so that differences in brain size, language ability, and "intelligence" would seem to be the consequences of a small set of genetic changes. With advances in the technical means for sequencing nucleic acids and with the accumulation of knowledge about the gene sequences of humans, chimpanzees, and other mammals from the various genome projects, we are now in a position to identify the genetic specializations of human beings. These specializations turn out to be far more extensive than expected, and include not only changes in gene sequences and gene expression, but also rearrangements, duplications, and losses of blocks of DNA.

Many of the gene-sequence differences that distinguish humans from other animals are the result of random substitutions (genetic drift), so it is necessary to cull these from the list of sequence differences to identify those that were likely to have been produced by natural selection. The culling is done statistically: genes under selection are expected to have ratios of nonsynonymous-to-synonymous sequence changes (i.e., nucleotide substitutions that produce amino-acid substitutions versus those that do not) greater than expected by chance. Among the genes that appear to have undergone positive selection in the human lineage are those for *FOXP2* (forkhead box P2) (Enard, Przeworski, et al., 2002), *MCH1* (microcephalin) (Evans et al., 2005), and *ASPM* (abnormal spindlelike microcephaly associated) (Mekel-Bobrov et al., 2005). *FOXP2* is of special interest because mutations of the gene are associated with language disorder, while specific mutations of *MCH1* and *ASPM* are associated with microcephaly. The suggestion, then, is that selection on these genes is related to the evolution of language (*FOXP2*) and increased brain size (*MCH1*, *ASPM*). Complementary to these single-gene studies, geneticists have also used high-throughput approaches (i.e., bioinformatic methods for comparing thousands of genes in genomic databases), identifying hundreds of DNA elements likely to have undergone sequence changes as the result of positive or negative selection (Arbiza, Dopazo, & Dopazo, 2006; Bustamante et al., 2005; Clark et al., 2003; Harris & Meyer, 2006). Some of selection-modified sequences are protein-coding genes, while others are regulatory elements. Relaxation of selection pressure has also been a factor in human genetic evolution, resulting, for example, in the transformation of numerous olfactory receptor genes into nonfunctional pseudogenes (reviewed by Roquier & Giorgi, 2007).

King and Wilson (1975), noting the ~1% difference in protein-coding gene sequences between humans and chimpanzees, suggested that the anatomical and behavioral differences between these species are mainly the result of evolutionary changes in gene expression. Whole-genome sequencing projects have paved the way for comparative studies of gene expression by providing the information required to make gene microarrays. Microarrays are comprised of probes representing a significant fraction of the genes in the genome. Messenger RNA extracted from tissue is labeled with a fluorescent marker, and the amount of mRNA bound to specific probe sequences provides an index of the strength of expression of specific genes in the tissue sample. In this way, gene expression in homologous tissues can be compared across species. A number of such studies have now been carried out (Cáceres et al., 2003; Enard, Khaitovich, et al., 2002; Khaitovich et al., 2005; Uddin et al., 2004; see also Gu & Gu, 2003; Hsieh, Chu, Wolfinger, & Gibson, 2003), and, as with the high-throughput sequence studies, they suggest that hundreds of genes underwent expression changes in human evolution (Preuss et al., 2004). Evolutionary changes in promoter sites and transcription factors that could account for these expression changes are areas of active investigation (e.g., Donaldson & Gottgens, 2006; Hammock & Young, 2005; Heissig et al., 2005; Pollard, Salama, King, et al., 2006; Rockman et al., 2005; Spiteri et al., 2007; Vernes et al., 2007). Understanding the functional significance of the changes has been challenging, owing in part to the variety of functional classes of genes involved. Recently, Oldham, Horvath, and Geschwind (2006) addressed this issue using network-analytic approach. They identified several sets of gene-coexpression "modules" common to humans and chimpanzees, along with a cortical module comprising multiple genes involved in energy metabolism that was present in humans but nearly absent in chimpanzees. Related to the latter module also were a number of genes involved in synapse formation and function.

These microarray studies have several limitations. One is that the microarrays used were constructed with gene sequences from humans, and sequence differences between humans and the other species examined bias the results (Hsieh et al., 2003; Preuss et al., 2004). Human-chimpanzee sequence differences can be corrected for, but comparisons with more distantly related species are problematic. New techniques for quantifying gene expression using mass resequencing should overcome these problems. Another limitation is that expression studies to date have been restricted to comparisons of adult brain tissues. Presumably, some of the most important differences between humans and nonhuman primates, including increased brain size, reflect changes in gene expression occurring in early development. Identifying these will require comparing fetal tissues from humans and apes, and availability of needed materials is problematic.

Identifying sequences and expression changes is just the first step in relating genetic to phenotypic changes:

if we want to know how genetic changes are affecting human phenotypes, we have to determine where and how those genes are acting in humans compared to other primates. This is where expression studies have an advantage over sequence studies. In sequence studies, one can know that a particular gene was selected for in human evolution, but since most genes are expressed in multiple tissues and cell types, one does not know which tissue or tissues were the targets of selection. Expression studies compare RNA samples from homologous tissues from different species, so they can provide starting points for comparative studies of those tissues. In one of the first attempts to do so, Cáceres and colleagues (Cáceres, Suwyn, Maddox, Thomas, & Preuss, 2007) compared the expression of *THBS4* (thrombospondin 4) in frontal cortex of humans, chimps, and macaques. Experimental studies in rodents and cell culture suggest that thrombospondin proteins stimulate neurons to make synapses (Christopherson et al., 2005; Susman et al., 2007), and thus increased expression in human cortex suggests evolutionary changes in synaptic turnover or plasticity. Thrombospondins are, however, expressed by a variety of brain cell types (neurons, glia, endothelial cells) and mediate many functions in addition to synaptogenesis. Cáceres and colleagues examined the expression of THBS4 protein in frontopolar cortex, and found the THBS4 protein is distributed much more densely within the cortical neuropil (where synapses are concentrated) of humans than in other species (figure 3.5). This finding is consistent with a link between changes in thrombospondin expression and synaptic dynamics in human evolution, but it is far from proof. Additional study is required, but now we have a good idea where to look. Through studies like these, we can use our rapidly growing knowledge of changes in any gene sequence or expression level to drive "phenotype discovery" at the cell and tissue levels (Preuss et al., 2004).

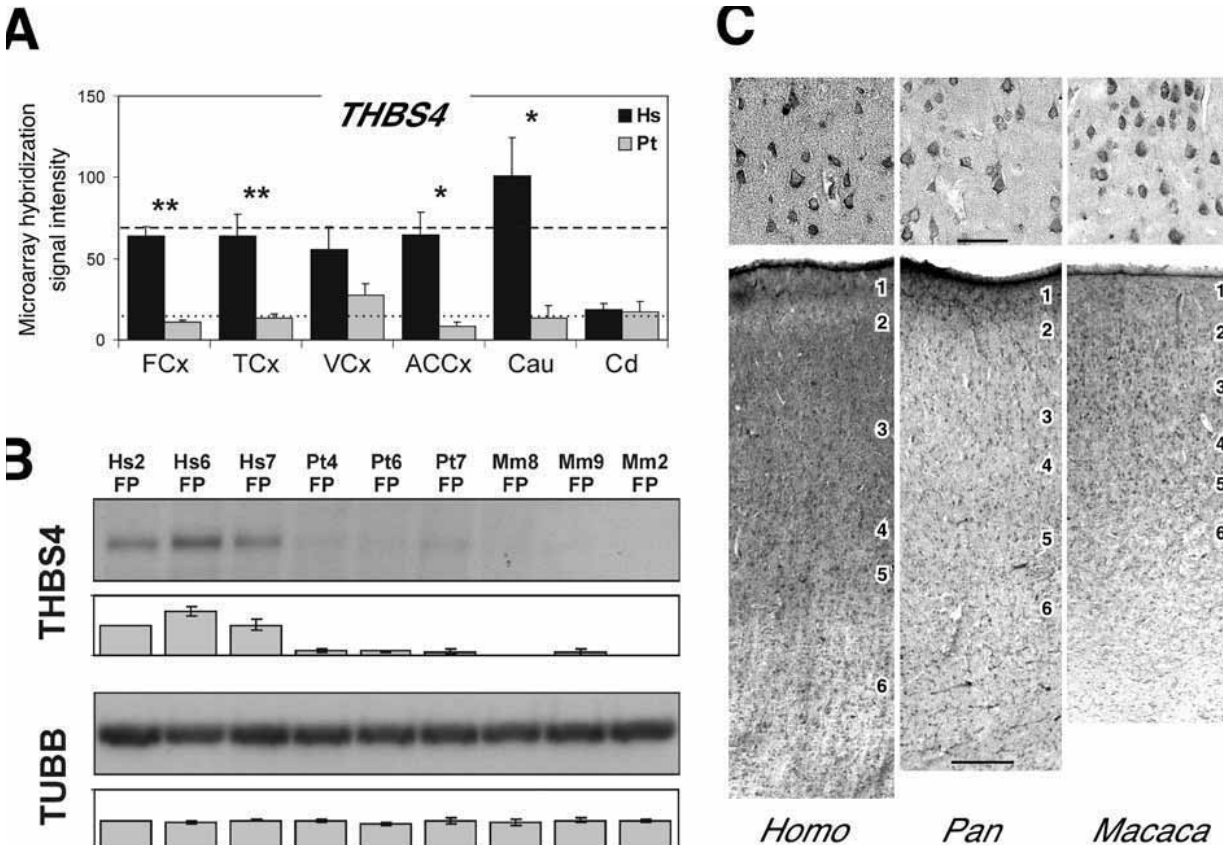


FIGURE 3.5 Increased expression of the thrombospondin 4 (*THBS4*) gene and protein in human brain evolution, according to Cáceres, Suwyn, Maddox, Thomas, and Preuss (2007). (*A*) Microarray analysis of *THBS4* mRNA levels in different brain regions of humans (Hs, *Homo sapiens*) and chimpanzees (Pt, *Pan troglodytes*). Expression levels were significantly higher in human frontal cortex (FCx), temporal cortex (TCx), anterior cingulate cortex (ACCx), and caudate (Cd). (*B*) Western blots showing higher levels of THBS4 protein levels in frontal cortex from three individual humans (Hs), compared to three chimpanzees (Pt) and three rhesus macaques (*Macaca mulatta*, Mm). Tubulin (TUBB) served as a loading control. (*C*) Immunocytochemistry for THBS4 in sections for frontopolar cortex yielded much denser labeling of humans than of chimpanzees or macaques. The difference was especially strong in the neuropil space surrounding neuronal somas. Scales are 50 microns in the upper panel and 250 microns in the lower panel.

Phenotype discovery would not be an important matter if it really were the case that the differences between humans and chimpanzees come down to just a few key genes. Recent studies, however, indicate that the genetic differences between humans and chimpanzees have been substantially understated (Cohen, 2007c), and the actual difference in total DNA sequence is closer to 4% than 1% (Britten, 2002; Varki & Altheide, 2005). The increase reflects the discovery that multiple duplications and deletions of DNA blocks occurred independently in human and chimpanzee evolution (J. Bailey & Eichler, 2006; Cheng et al., 2005; Sikela, 2006). This implies, perhaps surprisingly, that humans have genes that chimpanzees do not have, and vice versa (Eichler et al., 2001; Hayakawa et al., 2005; Johnson et al., 2001; Popesco et al., 2006; Sikela, 2006; Varki & Altheide, 2005). Moreover, these differences are just part of a much larger set of macromolecular difference between humans and nonhuman primates. For example, humans and nonhuman primates have different complements of expressed, but untranslated, RNAs, molecules that are thought to regulate mRNA translation (Berezikov et al., 2006; Pollard, Salama, Lambert, et al., 2006; Zhang, Peng, Wang, & Su, 2007). Humans also exhibit specializations in the way the multiple domains of protein-coding genes are spliced to form expressed proteins (Calarco et al., 2007). Finally, there are human specializations of the posttranslational modifications of proteins, specializations that may be especially relevant to infectious and neurodegenerative disease (Brooks, 2004; Gearing, Tigges, Mori, & Mirra, 1996; Varki, 2006; Walker, Rosen, & LeVine, 2008). There is thus a myriad of macromolecular differences between humans and other species, the functional and phenotypic significance of which awaits elucidation. The large number of genetic differences between humans and chimpanzees is all the more impressive when one considers how little reliable empirical evidence we have about the psychological, neurobiological, and other phenotypic differences between humans and chimpanzees (e.g., Gagneux & Varki, 2001; Gibbs, Collard, & Wood, 2002; Varki et al., 1998).

Conclusions

As neuroscientists, the species we would most like to understand, and most need to understand, is our own. Our discipline, however, rests largely on a foundation of studies in nonhuman species, and a very small set of nonhuman species at that. That strategy might have seemed reasonable when one could seriously maintain that the important features of brains were basically the same across species and when we lacked the means to study the human brain directly in much detail. The situation has now changed. We know that evolution produced a diversity of brain organizations and that the human brain, like that of every other species, possesses species-specific features in addition to features shared with other animals. We have just begun the work of identifying the shared and distinctive features of the human brain. The powerful noninvasive neuroscientific methods that have recently become available provide us with the means to pursue this project, for not only do they make the inner workings of the human brain more accessible to us, but they also make it possible for us to compare humans and other species as never before.

The advent of our modern understanding of the evolutionary relationship between humans and other animals, and the development of these new investigative techniques, should prompt a reevaluation of neuroscientific research strategies and resource allocation. The science we have built, centered on the model-animal paradigm and supported by biomedical funding agencies, is in important respects the wrong kind of science for elucidating the structure, functions, and diseases of the human brain. If we want to understand how humans resemble and differ from other animals—a goal that is central not only to the broad intellectual program of neuroscience but also critical for advancing its biomedical goals—then we need rigorous comparative studies, studies that involve more species than just the anointed few models (Preuss, 2000a, 2006). I am not suggesting that we abandon our model species, but rather that those species are not enough.

If our goal is to understand humans, the nonhuman species we can least afford to do without are the great apes, and most especially chimpanzees. As the examples cited in this essay demonstrate, the strategy of comparing humans and chimpanzees, our closest relatives, is a critical component of any research program that would make specific and well-founded claims about the human brain or other aspects of human biology. Tragically, just at the point in time that we have acquired the technical means to carry out these comparisons, in the form of noninvasive neuroscience technologies, we are on the brink of losing the other essential resource: the chimpanzees. Wild chimpanzee populations are on a path to extinction. Nevertheless, the National Institutes of Health, which holds a large fraction of the chimpanzees resident in the United States, has decided to no longer support the propagation of their captive chimpanzees, resulting in the eventual elimination of its colonies (Cohen, 2007a, 2007b). Remaining zoo populations are likely too small to be sustainable. Unless we take rapid action to maintain these uniquely valuable animals and do so in a way that makes them accessible to benign, noninvasive research, our ability to understand what makes us human will be forever diminished.

ACKNOWLEDGMENTS The author is grateful to acknowledge the support of the James S. McDonnell Foundation (JSMF 21002093), the Yerkes National Primate Research Center under NIH/NRCC grant RR00165, and the Center for Behavioral Neuroscience under the STC program of the National Science Foundation (IBN-9876754).

REFERENCES

Allman, J. M. (1977). Evolution of the visual system in the early primates. In J. M. Sprague & A. N. Epstein (Eds.), *Progress in psychology and physiological psychology* (Vol. 7, pp. 1–53). New York: Academic Press.

Allman, J. M. (1999). *Evolving brains*. New York: Scientific American Library.

Allman, J. M., Watson, K. K., Tetreault, N. A., & Hakeem, A. Y. (2005). Intuition and autism: A possible role for Von Economo neurons. *Trends Cogn. Sci.*, *9*(8), 367–373.

Arbiza, L., Dopazo, J., & Dopazo, H. (2006). Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. *PLoS Comput. Biol.*, *2*(4), e38.

Bailey, J. A., & Eichler, E. E. (2006). Primate segmental duplications: Crucibles of evolution, diversity and disease. *Nature Rev. Genet.*, *7*(7), 552–564.

Bailey, P., Bonin, G. v., & McCulloch, W. S. (1950). *The isocortex of the chimpanzee*. Urbana: University of Illinois Press.

Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, *34*(1), 144–155.

Benzing, W. C., Kordower, J. H., & Mufson, E. J. (1993). Galanin immunoreactivity within the primate basal forebrain: Evolutionary change between monkeys and apes. *J. Comp. Neurol.*, *336*(1), 31–39.

Berezikov, E., Thuemmler, F., van Laake, L. W., Kondova, I., Bontrop, R., Cuppen, E., et al. (2006). Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.*, *38*(12), 1375–1377.

Britten, R. J. (2002). Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl. Acad. Sci. USA*, *99*(21), 13633–13635.

Brodmann, K. (1909). *Lokalisationslehre der Grosshirnrhinde*. Leipzig: Barth. (Reprinted as *Brodmann's "Localisation in the cerebral cortex,"* trans. & ed. L. J. Garey, London: Smith-Gordon, 1994.)

Brooks, S. A. (2004). Appropriate glycosylation of recombinant proteins for human use: Implications of choice of expression system. *Mol. Biotechnol.*, *28*(3), 241–255.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn. Sci.*, *11*(2), 49–57.

Buckner, R. L., & Vincent, J. L. (2007). Unrest at rest: Default activity and spontaneous network correlations. *Neuroimage*, *37*(4), 1091–1096; discussion, 1097–1099.

Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M. T., Glanowski, S., et al. (2005). Natural selection on protein-coding genes in the human genome. *Nature*, *437*(7062), 1153–1157.

Buxhoeveden, D. P., Switala, A. E., Litaker, M., Roy, E., & Casanova, M. F. (2001). Lateralization of minicolumns in human planum temporale is absent in nonhuman primate cortex. *Brain Behav. Evol.*, *57*(6), 349–358.

Cáceres, M., Lachuer, J., Zapala, M. A., Redmond, J., Kudo, L., Geschwind, D., et al. (2003). Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci. USA*, *100*, 1330–1335.

Cáceres, M., Suwyn, C., Maddox, M., Thomas, J. W., & Preuss, T. M. (2007). Increased cortical expression of two synaptogenic thrombospondins in human brain evolution. *Cereb. Cortex*, *17*(19), 2312–2321 [Epub 2006 Dec 2320].

Calarco, J. A., Xing, Y., Cáceres, M., Calarco, J. P., Xiao, X., Pan, Q., et al. (2007). Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.*, *21*(22), 2963–2975.

Cantallops, I., & Routtenberg, A. (2000). Kainic acid induction of mossy fiber sprouting: Dependence on mouse strain. *Hippocampus*, *10*(3), 269–273.

Cheng, Z., Ventura, M., She, X., Khaitovich, P., Graves, T., Osoegawa, K., et al. (2005). A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature*, *437*(7055), 88–93.

Christopherson, K. S., Ullian, E. M., Stokes, C. C., Mullowney, C. E., Hell, J. W., Agah, A., et al. (2005). Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. *Cell*, *120*(3), 421–433.

Clark, A. G., Glanowski, S., Nielsen, R., Thomas, P. D., Kejariwal, A., Todd, M. A., et al. (2003). Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science*, *302*(5652), 1960–1963.

Cohen, J. (2007a). Animal studies: NIH to end chimp breeding for research. *Science*, *316*(5829), 1265.

Cohen, J. (2007b). Biomedical research: The endangered lab chimp. *Science*, *315*(5811), 450–452.

Cohen, J. (2007c). Evolutionary biology. Relative differences: The myth of 1%. *Science*, *316*(5833), 1836.

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., et al. (1999). Tracking neuronal fiber pathways in the living human brain. *Proc. Natl. Acad. Sci. USA*, *96*(18), 10422–10427.

Corballis, M. C. (2007). The evolution of hemispheric specializations of the human brain. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems:* Vol. 4. *Primates* (pp. 379–394). Oxford, UK: Elsevier.

Creutzfeldt, O. D. (1977). Generality of functional structure of the neocortex. *Naturwissenschaften*, *64*, 507–517.

Crick, F., & Jones, E. (1993). Backwardness of human neuroanatomy. *Nature*, *361*(6408), 109–110.

Croxson, P. L., Johansen-Berg, H., Behrens, T. E., Robson, M. D., Pinsk, M. A., Gross, C. G., et al. (2005). Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *J. Neurosci.*, *25*(39), 8854–8866.

Darwin, C. (1859). *On the origin of species*. London: John Murray. (Facsimile of first edition: Cambridge, MA: Harvard University Press, 1984.)

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: John Murray. (Facsimile edition: Princeton, NJ: Princeton University Press, 1981.)

Dauguet, J., Peled, S., Berezovskii, V., Delzescaux, T., Warfield, S. K., Born, R., et al. (2007). Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain. *Neuroimage*, *37*(2), 530–538.

DeFelipe, J., Alonso-Nanclares, L., & Arellano, J. I. (2002). Microstructure of the neocortex: Comparative aspects. *J. Neurocytol.*, *31*(3–5), 299–316.

DeFelipe, J., Alonso-Nanclares, L., Arellano, J., Ballesteros-Yáñez, I., Benavides-Piccione, R., & Muñoz, A. (2007). Specializations of the cortical microstructure of humans. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems:* Vol. 4. *Primates* (pp. 167–190). Oxford, UK: Elsevier.

Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., Van Essen, D., et al. (2004). The processing of visual shape in the cerebral cortex of human and nonhuman primates:

A functional magnetic resonance imaging study. *J. Neurosci.*, *24*(10), 2551–2565.

De Sousa, A., & Wood, B. (2007). The hominin fossil record and the emergence of the modern human central nervous system. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems:* Vol. 4. *Primates* (pp. 291–336). Oxford, UK: Elsevier.

Donaldson, I. J., & Gottgens, B. (2006). Evolution of candidate transcriptional regulatory motifs since the human-chimpanzee divergence. *Genome Biol.*, *7*(6), R52.

Eichler, E. E., Johnson, M. E., Alkan, C., Tuzun, E., Sahinalp, C., Misceo, D., et al. (2001). Divergent origins and concerted expansion of two segmental duplications on chromosome 16. *J. Hered.*, *92*(6), 462–468.

Elliot Smith, G. (1924). *The evolution of man: Essays*. London: Oxford University Press.

Elston, G. N. (2007). Specialization of the neocortical pyramidal cell during primate evolution. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems:* Vol. 4. *Primates* (pp. 191–242). Oxford, UK: Elsevier.

Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., et al. (2002). Intra- and interspecific variation in primate gene expression patterns. *Science*, *296*(5566), 340–343.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*(6900), 869–872.

Evans, P. D., Gilbert, S. L., Mekel-Bobrov, N., Vallender, E. J., Anderson, J. R., Vaez-Azizi, L. M., et al. (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science*, *309*(5741), 1717–1720.

Fleagle, J. G. (1999). *Primate adaptation and evolution*, 2nd ed. San Diego: Academic Press.

Gagneux, P., & Varki, A. (2001). Genetic differences between humans and great apes. *Mol. Phylogenet. Evol.*, *18*(1), 2–13.

Gannon, P. J., Holloway, R. L., Broadfield, D. C., & Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: Humanlike pattern of Wernicke's brain language area homolog. *Science*, *279*, 220–222.

Gearing, M., Tigges, J., Mori, H., & Mirra, S. S. (1996). A beta40 is a major form of beta-amyloid in nonhuman primates. *Neurobiol. Aging*, *17*(6), 903–908.

Geschwind, N. (1970). The organization of language and the brain. *Science*, *170*(961), 940–944.

Geschwind, N., & Levitsky, W. (1968). Left-right asymmetry in temporal speech region. *Science*, *161*, 186–187.

Gibbs, S., Collard, M., & Wood, B. (2002). Soft-tissue anatomy of the extant hominoids: A review and phylogenetic analysis. *J. Anat.*, *200*(Pt. 1), 3–49.

Gilissen, E. (2001). Structural symmetries and asymmetries in human and chimpanzee brains. In D. Falk & K. R. Gibson (Eds.), *Evolutionary anatomy of the primate cerebral cortex* (pp. 187–215). Cambridge, UK: Cambridge University Press.

Gingerich, P. D., Haq, M., Zalmout, I. S., Khan, I. H., & Malkani, M. S. (2001). Origin of whales from early artiodactyls: Hands and feet of Eocene Protocetidae from Pakistan. *Science*, *293*(5538), 2239–2242.

Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cereb. Cortex*, *18*(11), 2471–2482.

Gu, J., & Gu, X. (2003). Induced gene expression in human brain after the split from chimpanzee. *Trends Genet.*, *19*(2), 63–65.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nat. Rev. Neurosci.*, *2*(10), 685–694.

Haeckel, E. (1874). *Anthropogenie, oder, Entwickelungsgeschichte des Menschen*. Leipzig: W. Engelmann.

Haeckel, E. (1879). *The evolution of man: A popular exposition of the principal points of human ontogeny and phylogeny*. New York: Appleton.

Hammock, E. A., & Young, L. J. (2005). Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science*, *308*(5728), 1630–1634.

Harris, E. E., & Meyer, D. (2006). The molecular signature of selection underlying human adaptations. *Am. J. Phys. Anthropol.*, *Suppl. 43*, 89–130.

Hayakawa, T., Angata, T., Lewis, A. L., Mikkelsen, T. S., Varki, N. M., & Varki, A. (2005). A human-specific gene in microglia. *Science*, *309*(5741), 1693.

Heissig, F., Krause, J., Bryk, J., Khaitovich, P., Enard, W., & Paabo, S. (2005). Functional analysis of human and chimpanzee promoters. *Genome Biol.*, *6*(7), R57.

Hof, P. R., Nimchinsky, E. A., Perl, D. P., & Erwin, J. M. (2001). An unusual population of pyramidal neurons in the anterior cingulate cortex of hominids contains the calcium-binding protein calretinin. *Neurosci. Lett.*, *307*(3), 139–142.

Hof, P. R., & Sherwood, C. C. (2005). Morphomolecular neuronal phenotypes in the neocortex reflect phylogenetic relationships among certain mammalian orders. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.*, *287*(1), 1153–1163.

Hof, P. R., & Sherwood, C. C. (2007). The evolution of neuron classes in the neocortex of mammals. In J. H. Kaas & L. A. Krubitzer (Eds.), *Evolution of nervous systems:* Vol. 3: *Mammals* (pp. 113–124). Oxford, UK: Elsevier.

Hof, P. R., & Van Der Gucht, E. (2007). Structure of the cerebral cortex of the humpback whale, *Megaptera novaeangliae* (Cetacea, Mysticeti, Balaenopteridae). *Anat. Rec. (Hoboken)*, *290*(1), 1–31.

Hoffman, G. E., & Le, W. W. (2004). Just cool it! Cryoprotectant antifreeze in immunocytochemistry and in situ hybridization. *Peptides*, *25*(3), 425–431.

Hopkins, W. D., Marino, L., Rilling, J. K., & MacGregor, L. A. (1998). Planum temporale asymmetries in great apes as revealed by magnetic resonance imaging (MRI). *Neuroreport*, *9*(12), 2913–2918.

Hsieh, W. P., Chu, T. M., Wolfinger, R. D., & Gibson, G. (2003). Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics*, *165*(2), 747–757.

Huxley, T. H. (1863). *Evidence as to man's place in nature*. London: Williams and Norgate [Ann Arbor: University of Michigan, 1959].

Johnson, M. E., Viggiano, L., Bailey, J. A., Abdul-Rauf, M., Goodwin, G., Rocchi, M., et al. (2001). Positive selection of a gene family during the emergence of humans and African apes. *Nature*, *413*(6855), 514–519.

Kaas, J. H. (1977). Sensory representations in mammals. In G. S. Stent (Ed.), *Function and formation of neural systems* (pp. 65–80). Berlin: Dahlem Konferenzen.

Kaas, J. H. (2007). Reconstructing the organization of neocortex of the first mammals and subsequent modifications. In J. H. Kaas & L. A. Krubitzer (Eds.), *Evolution of nervous systems:* Vol. 3. *Mammals* (pp. 27–48). Oxford, UK: Elsevier.

Kaas, J. H., & Preuss, T. M. (2008). Human brain evolution. In L. R. Squire, D. Berg, F. E. Bloom, S. du Lac, A. Ghosh,

& N. C. Spizer (Eds.), *Fundamental neuroscience*, 3rd ed. (pp. 1019–1037). Amsterdam: Academic Press.

Kaufman, J. A., Ahrens, E. T., Laidlaw, D. H., Zhang, S., & Allman, J. M. (2005). Anatomical analysis of an aye-aye brain (*Daubentonia madagascariensis*, primates: Prosimii) combining histology, structural magnetic resonance imaging, and diffusion-tensor imaging. *Anat. Rec. A Discov. Mol. Cell Evol. Biol.*, *287*(1), 1026–1037.

Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., et al. (2005). Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, *309*(5742), 1850–1854.

King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science*, *188*(4184), 107–116.

Kordower, J. H., & Mufson, E. J. (1990). Galanin-like immunoreactivity within the primate basal forebrain: Differential staining patterns between humans and monkeys. *J. Comp. Neurol.*, *294*(2), 281–292.

Le Gros Clark, W. E. (1959). *The antecedents of man*. Edinburgh: Edinburgh University Press.

Logan, C. A. (2001). "Are Norway rats . . . things?": Diversity versus generality in the use of albino rats in experiments on development and sexuality. *J. Hist. Biol.*, *34*(2), 287–314.

Logan, C. A. (2002). Before there were standards: The role of test animals in the production of empirical generality in physiology. *J. Hist. Biol.*, *35*(2), 329–363.

Logan, C. A. (2005). The legacy of Adolf Meyer's comparative approach: Worcester rats and the strange birth of the animal model. *Integr. Physiol. Behav. Sci.*, *40*(4), 169–181.

Lovejoy, A. O. (1964). *The great chain of being: A study of the history of an idea*. Cambridge, MA: Harvard University Press.

MacLeod, C. E., Zilles, K., Schleicher, A., Rilling, J. K., & Gibson, K. R. (2003). Expansion of the neocerebellum in Hominoidea. *J. Hum. Evol.*, *44*(4), 401–429.

Marino, L., Rilling, J. K., Lin, S. K., & Ridgway, S. H. (2000). Relative volume of the cerebellum in dolphins and comparison with anthropoid primates. *Brain Behav. Evol.*, *56*(4), 204–211.

Marks, J. (2002). *What it means to be 98% chimpanzee: Apes, people, and their genes*. Berkeley: University of California Press.

McNamara, R. K., Namgung, U., & Routtenberg, A. (1996). Distinctions between hippocampus of mouse and rat: Protein F1/GAP-43 gene expression, promoter activity, and spatial memory. *Brain Res. Mol. Brain. Res.*, *40*(2), 177–187.

Mekel-Bobrov, N., Gilbert, S. L., Evans, P. D., Vallender, E. J., Anderson, J. R., Hudson, R. R., et al. (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. *Science*, *309*(5741), 1720–1722.

Mori, S., & van Zijl, P. C. (2002). Fiber tracking: Principles and strategies—A technical review. *NMR Biomed.*, *15*(7–8), 468–480.

Mori, S., & Zhang, J. (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron*, *51*(5), 527–539.

Mountcastle, V. B. (1978). An organizing principle for cerebral function: The unit module and the distributed system. In G. M. Edelman (Ed.), *The mindful brain* (pp. 7–50). Cambridge, MA: MIT Press.

Murphy, W. J., Eizirik, E., O'Brien, S. J., Madsen, O., Scally, M., Douady, C. J., et al. (2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, *294*(5550), 2348–2351.

Nimchinsky, E. A., Gilissen, E., Allman, J. M., Perl, D. P., Erwin, J. M., & Hof, P. R. (1999). A neuronal morphologic type unique to humans and great apes. *Proc. Natl. Acad. Sci. USA*, *96*(9), 5268–5273.

Oldham, M. C., Horvath, S., & Geschwind, D. H. (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc. Natl. Acad. Sci. USA*, *103*(47), 17973–17978.

Orban, G. A., Claeys, K., Nelissen, K., Smans, R., Sunaert, S., Todd, J. T., et al. (2006). Mapping the parietal cortex of human and nonhuman primates. *Neuropsychologia*, *44*(13), 2647–2667.

Orban, G. A., Fize, D., Peuskens, H., Denys, K., Nelissen, K., Sunaert, S., et al. (2003). Similarities and differences in motion processing between the human and macaque brain: Evidence from fMRI. *Neuropsychologia*, *41*(13), 1757–1768.

Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn. Sci.*, *8*(7), 315–324.

Parker, G. J., Stephan, K. E., Barker, G. J., Rowe, J. B., MacManus, D. G., Wheeler-Kingshott, C. A., et al. (2002). Initial demonstration of in vivo tracing of axonal projections in the macaque brain and comparison with the human brain using diffusion tensor imaging and fast marching tractography. *Neuroimage*, *15*(4), 797–809.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.*, *31*(2), 109–130; discussion, 130–178.

Petrides, M., & Pandya, D. N. (1994). Comparative architectonic analysis of the human and the macaque frontal cortex. In F. Booler & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 9, pp. 17–58). Amsterdam: Elsevier.

Phillips, C. G., Zeki, S., & Barlow, H. B. (1983). Localization of function in the cerebral cortex: Past, present and future. *Brain*, *107*, 328–361.

Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., et al. (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.*, *2*(10), e168.

Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M. A., Coppens, S., Pedersen, J. S., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, *443*(7108), 167–172.

Popesco, M. C., Maclaren, E. J., Hopkins, J., Dumas, L., Cox, M., Meltesen, L., et al. (2006). Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science*, *313*(5791), 1304–1307.

Preuss, T. M. (1993). The role of the neurosciences in primate evolutionary biology: Historical commentary and prospectus. In R. D. E. MacPhee (Ed.), *Primates and their relatives in phylogenetic perspective* (pp. 333–362). New York: Plenum Press.

Preuss, T. M. (1995). The argument from animals to humans in cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1227–1241). Cambridge, MA: MIT Press.

Preuss, T. M. (2000a). Taking the measure of diversity: Comparative alternatives to the model-animal paradigm in cortical neuroscience. *Brain Behav. Evol.*, *55*(6), 287–299.

Preuss, T. M. (2000b). What's human about the human brain? In *The new cognitive neurosciences*, 2nd ed. (pp. 1219–1234). Cambridge, MA: MIT Press.

Preuss, T. M. (2001). The discovery of cerebral diversity: An unwelcome scientific revolution. In D. Falk & K. Gibson (Eds.), *Evolutionary anatomy of the primate cerebral cortex* (pp. 138–164). Cambridge, UK: Cambridge University Press.

Preuss, T. M. (2004a). Specializations of the human visual system: The monkey model meets human reality. In J. H. Kaas & C. E. Collins (Eds.), *The primate visual system* (pp. 231–259). Boca Raton, FL: CRC Press.

Preuss, T. M. (2004b). What is it like to be a human? In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (3rd ed., pp. 5–22). Cambridge, MA: MIT Press.

Preuss, T. M. (2006). Who's afraid of *Homo sapiens*? *J. Biomed. Discov. Collaboration*, *1*, 17.

Preuss, T. M. (2007a). Evolutionary specializations of primate brain systems. In M. J. Ravosa & M. Dagosto (Eds.), *Primate origins: Evolution and adaptations* (pp. 625–675). New York: Springer.

Preuss, T. M. (2007b). Primate brain evolution in phylogenetic context. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems:* Vol. 4. *Primates* (pp. 1–34). Oxford, UK: Elsevier.

Preuss, T. M., Cáceres, M., Oldham, M. C., & Geschwind, D. H. (2004). Human brain evolution: Insights from microarrays. *Nat. Rev. Genet.*, *5*(11), 850–860.

Preuss, T. M., & Coleman, G. Q. (2002). Human-specific organization of primary visual cortex: Alternating compartments of dense Cat-301 and calbindin immunoreactivity in layer 4A. *Cereb. Cortex*, *12*(7), 671–691.

Preuss, T. M., Qi, H., & Kaas, J. H. (1999). Distinctive compartmental organization of human primary visual cortex. *Proc. Natl. Acad. Sci. USA*, *96*(20), 11601–11606.

Raghanti, M. A., Stimpson, C. D., Marcinkiewicz, J. L., Erwin, J. M., Hof, P. R., & Sherwood, C. C. (2008a). Cholinergic innervation of the frontal cortex: Differences among humans, chimpanzees, and macaque monkeys. *J. Comp. Neurol.*, *506*(3), 409–424.

Raghanti, M. A., Stimpson, C. D., Marcinkiewicz, J. L., Erwin, J. M., Hof, P. R., & Sherwood, C. C. (2008b). Differences in cortical serotonergic innervation among humans, chimpanzees, and macaque monkeys: A comparative study. *Cereb. Cortex*, *18*(3), 584–597.

Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *Neuroimage*, *37*(4), 1083–1090; discussion, 1097–1089.

Ramnani, N., Behrens, T. E., Johansen-Berg, H., Richter, M. C., Pinsk, M. A., Andersson, J. L., et al. (2006). The evolution of prefrontal inputs to the cortico-pontine system: Diffusion imaging evidence from macaque monkeys and humans. *Cereb. Cortex*, *16*(6), 811–818.

Ramnani, N., Behrens, T. E., Penny, W., & Matthews, P. M. (2004). New approaches for exploring anatomical and functional connectivity in the human brain. *Biol. Psychiatry*, *56*(9), 613–619.

Rekart, J. L., Sandoval, C. J., & Routtenberg, A. (2007). Learning-induced axonal remodeling: Evolutionary divergence and conservation of two components of the mossy fiber system within Rodentia. *Neurobiol. Learn. Mem.*, *87*(2), 225–235.

Richards, R. J. (1987). *Darwin and the emergence of evolutionary theories of mind and behavior*. Chicago: University of Chicago Press.

Richards, R. J. (1992). *The meaning of evolution: The morphological construction and ideological reconstruction of Darwin's theory*. Chicago: University of Chicago Press.

Rilling, J. K. (2006). Human and nonhuman primate brains: Are they allometrically scaled versions of the same design? *Evol. Anthropol.*, *15*(2), 65–77.

Rilling, J. K., Barks, S. K., Parr, L. A., Preuss, T. M., Faber, T. L., Pagnoni, G., et al. (2007). A comparison of resting-state brain activity in humans and chimpanzees. *Proc. Natl. Acad. Sci. USA*, *104*(43), 17146–17151.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.*, *11*(4), 426–428.

Rilling, J. K., & Insel, T. R. (1999a). Differential expansion of neural projection systems in primate brain evolution. *NeuroReport*, *10*(7), 1453–1459.

Rilling, J. K., & Insel, T. R. (1999b). The primate neocortex in comparative perspective using magnetic resonance imaging. *J. Hum. Evol.*, *37*(2), 191–223.

Rilling, J. K., & Seligman, R. A. (2002). A quantitative morphometric comparative analysis of the primate temporal lobe. *J. Hum. Evol.*, *42*(5), 505–533.

Rockel, A. J., Hiorns, R. W., & Powell, T. P. S. (1980). The basic uniformity of structure of the neocortex. *Brain*, *103*, 221–224.

Rockman, M. V., Hahn, M. W., Soranzo, N., Zimprich, F., Goldstein, D. B., & Wray, G. A. (2005). Ancient and recent positive selection transformed opioid cis-regulation in humans. *PLoS Biol.*, *3*(12), e387.

Roquier, S., & Giorgi, D. (2007). The loss of olfactory receptor genes in human evolution. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems:* Vol. 4. *Primates* (pp. 129–139). Oxford, UK: Elsevier.

Schenker, N. M., Desgouttes, A. M., & Semendeferi, K. (2005). Neural connectivity and cortical substrates of cognition in hominoids. *J. Hum. Evol.*, *49*(5), 547–569.

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., et al. (2007). Association fibre pathways of the brain: Parallel observations from diffusion spectrum imaging and autoradiography. *Brain*, *130*(Pt. 3), 630–653.

Schmitz, C., & Hof, P. R. (2005). Design-based stereology in neuroscience. *Neuroscience*, *130*(4), 813–831.

Schoenemann, P. T., Sheehan, M. J., & Glotzer, L. D. (2005). Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nat. Neurosci.*, *8*(2), 242–252.

Semendeferi, K., & Damasio, H. (2000). The brain and its main anatomical subdivisions in living hominoids using magnetic resonance imaging. *J. Hum. Evol.*, *38*(2), 317–332.

Semendeferi, K., Damasio, H., Frank, R., & Van Hoesen, G. W. (1997). The evolution of the frontal lobes: A volumetric analysis based on three-dimensional reconstructions of magnetic resonance scans of human and ape brains. *J. Hum. Evol.*, *32*(4), 375–388.

Semendeferi, K., Lu, A., Schenker, N., & Damasio, H. (2002). Humans and great apes share a large frontal cortex. *Nat. Neurosci.*, *5*(3), 272–276.

Sereno, M. I., & Tootell, R. B. (2005). From monkeys to humans: What do we now know about brain homologies? *Curr. Opin. Neurobiol.*, *15*(2), 135–144.

Sherwood, C., & Hof, P. (2007). The evolution of neuron types and cortical histology in apes and humans. In J. Kaas & T. Preuss (Eds.), *Evolution of nervous systems:* Vol. 4. *Primates* (pp. 355–378). Oxford, UK: Elsevier.

Sikela, J. M. (2006). The jewels of our genome: The search for the genomic changes underlying the evolutionarily unique capacities of the human brain. *PLoS Genet.*, *2*(5), e80.

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., et al. (2007). Identification of the transcriptional

targets of FOXP2, a gene linked to speech and language, in developing human brain. *Am. J. Hum. Genet.*, *81*(6), 1144–1157.

Striedter, G. F. (2005). *Principles of brain evolution*. Sunderland, MA: Sinauer Associates.

Subiaul, F., Barth, J., Okamoto-Barth, S., & Povinelli, D. J. (2007). Human cognitive specializations. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems:* Vol. 4. *Primates* (pp. 509–528). Oxford, UK: Elsevier.

Susman, M. W., Eroglu, C., Chakraborty, C., Hubermann, A. D., Green, E. M., Annis, D., et al. (2007). Identification of the thrombospondin receptor that promotes CNS synaptogenesis. *Soc. Neurosci. Abstracts*, *571.2*.

Szentagothai, J. (1975). The "module-concept" in cerebral cortex architecture. *Brain Res.*, *95*(2–3), 475–496.

Thewissen, J. G., Williams, E. M., Roe, L. J., & Hussain, S. T. (2001). Skeletons of terrestrial cetaceans and the relationship of whales to artiodactyls. *Nature*, *413*(6853), 277–281.

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., et al. (1997). Functional analysis of V3A and related areas in human visual cortex. *J. Neurosci.*, *17*(18), 7060–7078.

Uddin, M., Wildman, D. E., Liu, G., Xu, W., Johnson, R. M., Hof, P. R., et al. (2004). Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc. Natl. Acad. Sci. USA*, *101*(9), 2957–2962.

Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., Todd, J. T., et al. (2002). Extracting 3D from motion: Differences in human and monkey intraparietal cortex. *Science*, *298*, 413–415.

Varki, A. (2006). Nothing in glycobiology makes sense, except in the light of evolution. *Cell*, *126*(5), 841–845.

Varki, A., & Altheide, T. K. (2005). Comparing the human and chimpanzee genomes: Searching for needles in a haystack. *Genome Res.*, *15*(12), 1746–1758.

Varki, A., Wills, C., Perlmutter, D., Woodruff, D., Gage, F., Moore, J., et al. (1998). Great ape phenome project? *Science*, *282*(5387), 239–240.

Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., Davies, K. E., et al. (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *Am. J. Hum. Genet.*, *81*(6), 1232–1250.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., et al. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, *447*(7140), 83–86.

Walker, L. C., Rosen, R. F., & Leffine, H., Ii. (2008). Diversity of Aβ deposits in the aged brain: A window on molecular heterogeneity? *Rom. J. Morphol. Embryol.*, *49*(1), 5–11.

Watson, K. K., & Allman, J. M. (2007). Role of spindle cells in the social cognition of apes and humans. In J. Kaas & T. Preuss (Eds.), *Evolution of nervous systems:* Vol. 4. *Primates* (pp. 479–484). Oxford, UK: Elsevier.

West, M. J. (1999). Stereological methods for estimating the total number of neurons and synapses: Issues of precision and bias. *Trends Neurosci.*, *22*(2), 51–61.

Zhang, R., Peng, Y., Wang, W., & Su, B. (2007). Rapid evolution of an X-linked microRNA cluster in primates. *Genome Res.*, *17*(5), 612–617.

4 Unraveling the Role of Neuronal Activity in the Formation of Eye-Specific Connections

LEO M. CHALUPA AND ANDREW D. HUBERMAN

ABSTRACT Since the pioneering studies of Wiesel and Hubel on the development and plasticity of ocular dominance columns in the visual cortex, it has been widely thought that correlated discharges of neighboring retinal ganglion cells play an instructive role in the formation of segregated eye-specific domains in the mammalian visual system. Here we review the relevant evidence and conclude that while correlated retinal discharges are required for the formation of segregated eye-specific projections in the visual cortex, there is reason to doubt that this is also the case at the level of the dorsal lateral geniculate. More likely, molecular cues play a key role in the stereotypic pattern of segregated retinogeniculate projections that characterize different species. As yet, the role of activity and the identity of the molecular cues involved in this process remain to be firmly established.

Since the latter half of the last century, the formation of eye-specific projections has served as a model system for exploring the development and plasticity of neural circuits. Early in development, retinal ganglion cell (RGC) projections to the dorsal lateral geniculate (dLGN) (Rakic, 1976; Linden, Guillery, & Cucchiaro, 1981; Shatz, 1983; Godement, Salaun, & Imbert, 1984) and dLGN projections to V1 (Hubel, Wiesel, & LeVay, 1977; LeVay, Stryker, & Shatz, 1978; LeVay, Wiesel, & Hubel, 1980; Rakic, 1976) are intermingled. Subsequently, they segregate into nonoverlapping eye-specific territories, and this process is believed to require neuronal activity. Indeed, the precise pattern of neural activity, as opposed to the mere presence of action potential, has been hypothesized to "instruct" the segregation process by engaging well-established synaptic plasticity mechanisms (Crair, 1999; Feller, 1999; Stellwagen & Shatz, 2002; Torborg, Hansen, & Feller, 2005).

Here, we provide a brief historical account of the experimental evidence that gave rise to the idea that patterned neural activity plays an instructive role in the formation of eye-specific visual projections. We then review recent studies that tested directly whether patterns of neural activity in fact provide the instructive cues required for the segregation of eye-specific projections to the dLGN. This chapter is an update of a chapter on this topic that we authored for the third edition of *The Cognitive Neurosciences* (Chalupa & Huberman, 2004). We have also offered re-cent reviews of the role of activity in the formation of eye-specific projections in other publications (Chalupa, 2007; Huberman, Feller, & Chapman, 2008).

Formation of ocular dominance columns

That neuronal activity has an influence on the development of visual system connections stems from the pioneering studies of Wiesel and Hubel. Their work showed that closure of one eye during a critical period in postnatal life rendered that eye permanently incapable of driving cortical cells (Wiesel & Hubel, 1965a, 1965b). Subsequently it was shown that this physiological effect was accompanied by a marked reduction in the amount of cortical territory innervated by geniculocortical axons representing the deprived eye and a dramatic expansion of the geniculocortical axons representing the nondeprived eye (Hubel et al., 1977; Shatz & Stryker, 1978). These deprivation studies demonstrated that activity-mediated competition between axons representing the two eyes allocates postsynaptic space in V1.

Transneuronal tracing of retinal-dLGN-V1 connections has been used to assess the formation of ocular dominance columns (ODCs) during early development. Monocular injections of transneuronal tracers indicated that axons representing the two eyes start out overlapped (Hubel et al., 1977; LeVay et al., 1978, 1980; Rakic, 1976) before gradually segregating into ODCs. This finding supported a role for retinal activity in ODC segregation in that abolishing action potentials in both eyes with intraocular injections of tetrodotoxin (TTX) prevented the emergence of ODCs in the visual cortex (Stryker & Harris, 1986). Thus

LEO M. CHALUPA Department of Ophthalmology and Vision Science, School of Medicine and Department of Neurobiology, Physiology and Behavior, College of Biological Sciences, University of California, Davis, California
ANDREW D. HUBERMAN Department of Neurobiology, Stanford University School of Medicine, Stanford, California

the generally accepted model was that, early in development, projections representing the two eyes overlapped in V1, and the formation of separate left- and right-eye ODCs reflected activity-mediated competition between the two eyes.

The notion that activity-mediated events drive the initial formation of ODCs was first challenged by Crowley and Katz (1999). These investigators removed both eyes of postnatal ferrets weeks before geniculocortical axons invade V1, and then after allowing the animals to survive into adulthood, injected anterograde tracers into one or the other putative eye-specific layer in the dLGN. Surprisingly, they found patches of label that resembled normal ODCs in these animals. Obviously, the formation of ODCs could not be due to retinal activity. The authors thus concluded that formation of ODCs "relies primarily on activity-independent cues, rather than on specific patterns of correlated activity" (Crowley & Katz, 1999). In a subsequent study, these investigators used the same tracing method to demonstrate that ODCs are present in the ferret cortex very shortly after geniculocortical axons innervate layer IV and are refractory to early imbalances in retinal activity (Crowley & Katz, 2000). The results of these studies implied that the "overlap" of eye-specific dLGN inputs to V1 seen in the earlier transneuronal experiments was spurious, reflecting leakage of transneuronal tracer into opposite-eye layers in the dLGN. The results of Crowley and Katz therefore raised doubt as to whether activity plays a role in the initial formation of ODCs.

Almost a decade after their publication, the studies of Crowley and Katz (1999, 2000) remain controversial. Some have argued that patches of label they observed in V1 may not represent ODCs because eye removal early in life alters dLGN layering (reviewed in Huberman et al., 2008). Moreover, a recent study has shown that altering spontaneous retinal activity in the first postnatal week permanently disrupts patterning of ODCs and binocular receptive fields in V1 of the ferret (Huberman, Speer, & Chapman, 2006). That study traced ODCs using transneuronal methods in adulthood, when spillover is less of a concern. Now, resolution of whether ODCs form precisely from the outset or through refinement will require developmental studies with tracers not prone to spillover that are capable of demonstrating eye-specific labeling patterns within layer IV of the visual cortex.

Formation of eye-specific inputs to the dLGN

Rakic (1976) pioneered the exploration of the prenatal visual system and was the first to show that in the embryonic macaque monkey the projections of the two eyes are initially overlapped in the dLGN before segregating into eye-specific layers. Subsequently, this was found to be the case in other species using a variety of tracing methods (mouse: Godement et al., 1984; ferret: Linden et al., 1981; Huberman, Stellwagen, & Chapman, 2002; Huberman et al., 2003; cat: Shatz, 1983). Rakic (1981) also showed that fetal monocular enucleation results in the maintenance of a widespread projection from the remaining eye to the dLGN and the visual cortex, demonstrating the importance of prenatal binocular interactions in this process. Similar results have been obtained following monocular enucleation in the fetal cat (Chalupa & Williams, 1984; Shook & Chalupa, 1986).

The cellular and molecular mechanisms underlying binocular competitive interactions were not directly addressed in these studies, but generalizing from the literature dealing with the plasticity of cortical ODCs, it was reasonable to assume that prenatal binocular competition is also activity mediated. To test this idea, Shatz and Stryker (1988) infused TTX into the brain of fetal cats before retinal segregation of retinogeniculate axons occurred. This procedure caused the projections of the two eyes to remain widespread and overlapping in the dLGN. An obvious implication of this study was that retinal ganglion cells are capable of discharging action potentials even before these cells can be activated by light.

This capability was demonstrated by recording action potentials from retinal ganglion cells in embryonic rats (Galli & Maffei, 1988). These experiments showed that neighboring ganglion cells exhibit correlated discharges (Maffei & Galli-Resta, 1990). Subsequent in vitro patch-clamp recordings documented the ontogeny of excitable membrane properties in retinal ganglion cells from the fetal cat (Skaliora, Scobey, & Chalupa, 1993; Skaliora, Robinson, Scobey, & Chalupa, 1995) and rat (Wang, Ratto, Bisti, & Chalupa, 1997). The introduction of in vitro multielectrode-array recordings (Meister, Wong, Baylor, & Shatz, 1991) and optical recording of intracellular calcium allowed for simultaneous recording of hundreds of retinal cells (Feller, Wellis, Stellwagen, Werblin, & Shatz, 1996; Wong, Meister, & Shatz, 1993). These studies revealed that the spontaneous bursts of correlated ganglion cell activity propagate across the developing retina in a wavelike manner. Such retinal waves have been now reported in many species, including chick, turtle, mouse, rabbit, rat, ferret, and cat (reviewed in Wong, 1999), suggesting that they are a ubiquitous feature of retinal development. Based on their specific spatial and temporal properties, retinal waves have been hypothesized to play an essential role in the segregation of left- and right-eye inputs to the dLGN (Feller, 1999; Sengpiel & Kind, 2002; Cohen-Corey, 2002), the establishment of retinotopic order in the superior colliculus (Butts & Rokshar, 2001), and the segregation of On and Off ganglion cell projections to the dLGN (Wong & Oakley, 1996). There is substantial evidence that waves are involved in retinotopic refinement

in the SC (McLaughlin, Torborg, Feller, & O'Leary, 2003), dLGN (Grubb, Rossi, Changeux, & Thompson, 2003), and V1 (Cang et al., 2005), and, as mentioned previously, normal waves are needed for ODC development (Huberman et al., 2006). Whether waves drive eye-specific segregation in the dLGN, however, remains controversial.

How might retinal waves drive eye-specific segregation? Eye-specific segregation is thought to occur because the correlated discharges of spatially adjacent ganglion cells in one eye are better capable of depolarizing neurons in the dLGN than the inputs stemming from the two eyes. In accordance with Hebbs postulate (Hebb, 1949), coactive inputs are preferentially stabilized relative to temporally uncorrelated inputs (Bi & Poo, 2001; Zhang & Poo, 2001).

Three lines of evidence have been invoked in support for this "fire together wire together" idea. First, the occurrence of waves coincides with the segregation of initially overlapping retinal projections (mouse: Muir-Robinson, Hwang, & Feller, 2002; ferret: Penn, Riquelme, Feller, & Shatz, 1998; Huberman et al., 2002, 2003). Second, abolishing all retinal activity prevents the formation of eye-specific projection patterns (Muir-Robinson et al.; Penn et al.; Huberman et al., 2002, 2003). Third, inducing an imbalance in the overall activity levels of the two eyes causes the more active eye to innervate more dLGN territory at the expense of the less active eye (Penn et al.; Stellwagen & Shatz, 2002).

Still, the direct testing of the idea that correlated retinal ganglion cell discharges per se are essential for the formation of eye-specific projections requires that one perturb the correlated activity of adjacent retinal ganglion cells without impacting the overall discharge levels of these neurons. The studies that blocked retinal waves did so by abolishing all ganglion-cell discharges (Penn et al., 1998; Muir-Robinson et al., 2002; Huberman et al., 2002), and the studies that altered the balance of waves in the two eyes either significantly increased or decreased the frequency of ganglion-cell action potentials (Penn et al.; Stellwagen & Shatz, 2002). Thus it has remained an open question whether the changes in eye-specific projections observed in these experiments were due to altering the pattern, as opposed to the level, of spontaneous retinal activity.

The first opportunity to address this issue was provided by the manufacture of a novel immunotoxin designed to selectively target cholinergic neurons (Gunhan, Choudary, Landerholm, & Chalupa, 2002). Although the cellular basis of retinal waves remains unclear, it is known that during the developmental phase when eye-specific segregation is occurring, spontaneous discharges of retinal ganglion cells are caused by acetylcholine released by starburst amacrine cells (Feller et al., 1996; Zhou & Zhou, 2000). Immunotoxin depletion of starburst amacrine cells should thus disrupt the correlated discharges of ganglion cells without completely blocking the activity of these neurons or significantly changing their overall levels of activity (Huberman et al., 2003).

This strategy was applied to newborn ferrets. In this species retinogeniculate projections are extensively overlapped at birth, segregating gradually to attain their adultlike state by postnatal day 10 (P10) (Linden et al., 1981; Penn et al., 1998; Huberman et al., 2002, 2003). Binocular injections of the immunotoxin were made on the day of birth (P0), and within 48 hours this procedure resulted in elimination of approximately 80% of the starburst amacrine cells. To assess the effects of this manipulation on the firing patterns of these neurons, whole-cell patch-clamp recordings were made from morphologically identified P2–P10 retinal ganglion cells. These recordings showed that ganglion cells were in fact spontaneously active in the cholinergic-cell-depleted retinas. However, their discharge properties were markedly aberrant; some cells fired spikes very infrequently compared to those in normal retinas, while others manifested a firing frequency that was much higher than normal. Importantly, while the discharges of individual cells were abnormal, the overall firing rate of the sample of cells recorded was not statistically different from normal in the immunotoxin treated retinas (see figure 3 in Huberman et al., 2003).

To test directly whether correlated discharges were perturbed by starburst amacrine cell depletion, dual patch-clamp recordings were made from neighboring ganglion cells (less than 25 μm apart). In the normal P2–P10 retinas, all ganglion cells showed highly correlated spiking and membrane potential activity, irrespective of cell class (Huberman et al., 2003; Liets, Olshausen, Wang, & Chalupa, 2003). By contrast, not a single ganglion cell pair in the immunotoxin treated retinas showed any significant correlated activity, and this was the case for every retina examined from P2 to P10. These electrophysiological findings demonstrated unequivocally that depletion of starburst amacrine cells rapidly eliminated the correlated discharges of neighboring retinal ganglion cells, without significantly altering the overall activity levels of these neurons.

Did elimination of correlated ganglion-cell activity disrupt the segregation of left- and right-eye retinogeniculate projections? To answer this question, we examined the pattern of retinogeniculate connections in normal developing ferrets age P2 and P10 and compared these to P10 animals that received intraocular injections of the cholinergic immunotoxin on P0. As mentioned earlier, an injection of immunotoxin on P0 eliminated the correlated firing of neighboring ganglion cells by P2, an age when binocular retinogeniculate inputs are still intermingled extensively. Remarkably, in ferrets injected with immunotoxin, normal eye-specific segregation of retinogeniculate connections still formed. Indeed, quantitative comparison of the extent of overlap for left- and right-eye axons indicated that the degree and pattern of segregation was the same for control and

experimental P10 animals. These results showed for the first time that correlated activity of neighboring ganglion cells is not required for the segregation of eye-specific retinogeniculate projections.

The ongoing quest for understanding the role of retinal waves

Using mutant mice in which the β2 subunit of the nicotinic acetylsholine receptor is eliminated, Torborg and colleagues (2005) showed that retinal waves as well as eye-specific dLGN segregation are both perturbed. They concluded that a specific parameter of retinal activity—the amount of RGC bursting > 10 Hz—is the key element that instructs eye-specific segregation. However, a recent study has shown that retinal waves are in fact present in β2 mutant mice (Chao, Warland, Ballesteros, van der List, & Chalupa, 2008). Indeed, these mutants manifest robust waves of retinal activity, but eye-specific inputs to the dLGN fail to form normally.

Another approach we employed to test the relevance of retinal waves to the formation of segregated retinogeniculate projections was to assess these two phenomena in the fetal macaque monkey. In the species commonly used for studies of the developing visual system, such as the ferret and even more so the mouse, the time period during which connections in the visual system are made and refined is relatively brief. By contrast, the developmental period when brain connections become established is much more protracted in the macaque monkey, so we reasoned that the salient features of visual system development could be more dispersed, making it more likely that key events could be temporally dissociated from each other.

As a first step we sought to accurately discern the time period during which eye-specific retino-dLGN projections are established in the fetal macaque monkey (Huberman, Dehay, Berland, Chalupa, & Kennedy, 2005). This involved making intraocular injections of different tracers in each eye of fetal monkeys of known gestational ages. Subsequently, the extent of the binocular overlap within the developing dLGN was quantified using confocal microscopy and image analysis software. This study showed that about 100 days before birth at embryonic day 69 (E69), the projections from the two eyes were extensively intermingled in the dLGN. At E84 segregation of left and right eye axons was found to be essentially complete, with the six eye-specific domains that characterize the mature macaque dLGN clearly apparent. Thus the segregation of eye-specific inputs occurs during a remarkably early and relatively brief in utero period in the macaque, taking about 14 days in a 165-day gestation period.

Next, we determined whether the developmental period when eye-specific projections are formed relates to the presence of retinal waves in the fetal macaque retina. To address this issue, we used multielectrode arrays to record the activity from isolated retinas obtained from fetal monkeys of known gestational age (Warland, Huberman, & Chalupa, 2006). Before E55 the fetal monkey retina was found to be essentially silent, with only a few cells manifesting occasional discharges. This finding means that the specific ingrowth of retinogeniculate axons into regions destined to form magnocellular and parvocellular layers, which occurs prior to E55 (Meissirel, Wikler, Chalupa, & Rakic, 1997), is very unlikely to depend on retinal activity. Retinal waves of activity were found to be prevalent at E60, which is more than a week before segregation of retinogeniculate projections is first observed. Moreover, the incidence of retinal waves decreased progressively during the period when eye-specific projections become established (E69–E76). These findings in the fetal macaque monkey differ from the results obtained in other species where retinal wave activity has been reported to be robust throughout the period when segregated retinogeniculate projections are formed, with such activity declining after this developmental event becomes established (mouse: Demas et al., 2003; Torborg et al., 2005; ferret: Meister et al., 1991; Wong, Meister, & Shatz, 1993; Feller et al., 1996). Thus, while the multiarray recordings from fetal monkey retina do not rule out the possibility that retinal activity plays a role in the formation of segregated retinogeniculate projections, they do indicate that retinal waves are most prominent at a much earlier time in development than the period when eye-specific inputs are being established.

What accounts for the stereotyped pattern of eye-specific layers in the dLGN? Such stereotypy is difficult to reconcile with an activity-based mechanism. Indeed, a purely activity-dependent sorting should result in randomly distributed patterns of right- and left-eye inputs to the dLGN (Huberman et al., 2002; Muir-Robinson et al., 2002). Thus some bias for one or the other layer by the two eyes must exist (Sanes & Yamagata, 1999). Activity-dependent proponents have speculated that the stereotyped pattern of layering in the dLGN reflects the earlier arrival of contralateral-eye versus ipsilateral-eye axons. The use of sensitive anatomical tracers indicates, however, that axons from both eyes are present throughout the dLGN from very early times (Penn et al., 1998; Huberman et al., 2002, 2003).

Two recent studies (one in mouse and one in ferret) have shown that ephrin-As shape the formation of eye-specific zones (Huberman, Murray, Warland, Feldheim, & Chapman, 2005; Pfeiffenberger et al., 2005; Pfeiffenberger, Yamada, & Feldheim, 2006). This makes sense given that ephrin-A receptors, EphAs, are expressed in a gradient across the nasal temporal axis of the retina and that each eye-specific layer receives input from retinal ganglion cells in the nasal or temporal-retina.

Studies aimed at identifying eye-specific molecules in the dLGN and the genes that act downstream of activity to regulate axon refinement are currently under way in several laboratories. A handful of such molecules have been identified, and interestingly, all of these are immune-system factors (Huh et al., 2000; Bjartmar et al., 2006; Stevens et al., 2007). Nevertheless, the issue of how retinal activity contributes to the establishment of eye-specific projections still awaits resolution.

ACKNOWLEDGMENTS This work was supported by grants from the National Eye Institute of the National Institutes of Health and from a Research to Prevent Blindness Award.

REFERENCES

Bi, G., & Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.*, *24*, 139–66.

Bjartmar, L., Huberman, A. D., Ullian, E. M., Renteria, R. C., Liu, X., Xu, W., et al. (2006). Neuronal pentraxins mediate synaptic refinement in the developing visual system. *J. Neurosci.*, *26*, 6269–6281.

Butts, D. A., & Rokshar, D. S. (2001). The information content of spontaneous retinal waves. *J. Neurosci.*, *21*, 961–973.

Cang, J., Renteria, R., Kaneko, M., Liu, X., Copenhagen, D., & Stryker, M. (2005). Development of precise maps in visual cortex requires patterned spontaneous activity in the retina. *Neuron*, *48*, 797–809.

Chalupa, L. M. (2007). A reassessment of the role of activity in the formation of eye-specific retinogeniculate projections. *Brain Res. Brain Res. Rev.*, *55*, 228–236.

Chalupa, L. M., & Huberman, A. D. (2004). A new perspective on the role of activity in the development of eye-specific retinogeniculate projections. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (3rd ed., pp. 85–92). Cambridge MA: MIT Press.

Chalupa, L. M., & Williams, R. W. (1984). Organization of the cat's lateral geniculate nucleus following interruption of prenatal binocular competition. *Hum. Neurobiol.*, *3*, 103–107.

Chao, S., Warland, D. K., Ballesteros, J., van der List, D., & Chalupa, L. M. (2008). Retinal waves in mice lacking the β2 subunit of the nicotinic acetylcholine receptor. *Proc. Natl. Acad. Sci. USA*, *105*(36), 13638–13643.

Cohen-Cory, S. (2002). The developing synapse: Construction and modulation of synaptic structures and circuits. *Science*, *298*, 770–776.

Crair, M. C. (1999). Neuronal activity during development: Permissive or instructive? *Curr. Opin. Neurobiol.*, *9*, 88–93.

Crowley, J. C., & Katz, L. C. (1999). Development of ocular dominance columns in the absence of retinal input. *Nat. Neurosci.*, *2*, 1125–1130.

Crowley, J. C., & Katz, L. C. (2000). Early development of ocular dominance columns. *Science*, *290*, 1321–1324.

Demas, J., Sagdullaev, B. T., Green, E., Jaubert-Miazza, L., McCall, M. A., Gregg, R. G., et al. (2006). Failure to maintain eye-specific segregation in nob, a mutant with abnormally patterned retinal activity. *Neuron*, *50*, 247–259.

Feller, M. B. (1999). Spontaneous correlated activity in developing neural circuits. *Neuron*, *22*, 653–656.

Feller, M. B., Wellis, D. P., Stellwagen, D., Werblin, F., & Shatz, C. J. (1996). Requirement for cholinergic synaptic transmission in the propagation of spontaneous retinal waves. *Science*, *272*, 1182–1187.

Galli, L., & Maffei, L. (1988). Spontaneous impulse activity of rat retinal ganglion cells in prenatal life. *Science*, *242*, 90–91.

Godement, P., Salaun, J., & Imbert, M. (1984). Prenatal and postnatal development of retinogeniculate and retinocollicular projections in the mouse. *J. Comp. Neurol.*, *230*, 552–575.

Grubb, M. S., Rossi, F. M., Changeux, J. P., & Thompson, I. D. (2003). Abnormal functional organization in the dorsal lateral geniculate nucleus of mice lacking the beta 2 subunit of the nicotinic acetylcholine receptor. *Neuron*, *40*, 1161–1172.

Gunhan, E., Choudary, P. V., Landerholm, T. E., & Chalupa, L. M. (2002). Depletion of cholinergic amacrine cells by a novel immunotoxin does not perturb the formation of segregated On and Off cone bipolar cell projections. *J. Neurosci.*, *22*: 2265–2273.

Hebb, D. O. (1949). *Organization of behavior: A neuropsychological theory*. New York: John Wiley and Sons.

Hubel, D. H., Wiesel, T. N., & LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *278*, 377–409.

Huberman, A. D., Dehay, C., Berland, M., Chalupa, L. M., & Kennedy, H. (2005). Early and rapid targeting of eye-specific axonal projections to the lateral geniculate nucleus in the fetal macaque. *J. Neurosci.*, *25*, 4014–4023.

Huberman, A. D., Feller, M. B., & Chapman, B. (2008). Mechanisms underlying development of visual maps and receptive fields. *Annu. Rev. Neurosci.*, *31*, 479–509.

Huberman, A. D., Murray, K. D., Warland, D. K., Feldheim, D. A., & Chapman, B. (2005). Ephrin-As mediate targeting of eye-specific projections to the lateral geniculate nucleus. *Nat. Neurosci.*, *8*, 1013–1021.

Huberman, A. D., Speer, C. M., & Chapman, B. (2006). Spontaneous retinal activity mediates development of ocular dominance columns and binocular receptive fields in v1. *Neuron*, *52*, 247–254.

Huberman, A. D., Stellwagen, D., & Chapman, B. (2002). Decoupling eye-specific segregation from lamination in the lateral geniculate nucleus. *J. Neurosci.*, *22*, 9419–9429.

Huberman, A. D., Wang, G. Y., Liets, L. C., Collins, O. A., Chapman, B., & Chalupa, L. M. (2003). Eye-specific retinogeniculate segregation independent of normal neuronal activity. *Science*, *300*, 994–998.

Huh, G. S., Boulanger, L. M., Du, H., Riquelme, P. A., Brotz, T. M., & Shatz, C. J. (2000). Functional requirement for class I MHC in CNS development and plasticity. *Science*, *290*, 2155–2159.

LeVay, S. M., Stryker, P., & Shatz, C. J. (1978). Ocular dominance columns and their development in layer IV of the cat's visual cortex: A quantitative study. *J. Comp. Neurol.*, *179*, 223–224.

LeVay, S. M., Wiesel, T. N., & Hubel, D. H. (1980). The development of ocular dominance columns in normal and visually deprived monkeys. *J. Comp. Neurol.*, *191*, 1–51.

Liets, L. C., Olshausen, B. A., Wang, G. Y., & Chalupa, L. M. (2003). Spontaneous activity of morphologically identified ganglion cells in the developing ferret retina. *J. Neurosci.*, *23*, 7343–7350.

Linden, D. C., Guillery, R. W., Cucchiaro, J. (1981). The dorsal lateral geniculate nucleus of the normal ferret and its postnatal development. *J. Comp. Neurol.*, *203*, 189–211.

Maffei, L., & Galli-Resta, L. (1990). Correlation in the discharges of neighboring rat retinal ganglion cells during prenatal life. *Proc. Natl. Acad. Sci. USA*, *87*, 2861–2864.

McLaughlin, T., Torborg, C. L., Feller, M. B., & O'Leary, D. D. (2003). Retinotopic map refinement requires spontaneous retinal waves during a brief critical period of development. *Neuron*, *40*, 1147–1160.

Meissirel, C., Wikler, K. C., Chalupa, L. M., & Rakic, P. (1997). Early divergence of functional subsystems in the embryonic primate visual system. *Proc. Natl. Acad. Sci. USA*, *94*, 5900–5905.

Meister, M., Wong, R. O., Baylor, D. A., & Shatz, C. J. (1991). Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science*, *252*, 939–943.

Muir-Robinson, G., Hwang, B. J., & Feller, M. B. (2002). Retinogeniculate axons undergo eye-specific segregation in the absence of eye-specific layers. *J. Neurosci.*, *22*, 5259–5264.

Penn, A. A., Riquelme, P. A., Feller, M. B., & Shatz, C. J. (1998). Competition in retinogeniculate patterning driven by spontaneous activity. *Science*, *279*, 2108–2112.

Pfeiffenberger, C., Cutforth, T., Woods, G., Yamada, J., Renteria, R. C., Copenhagen, D. R., et al. (2005). Ephrin-As and neural activity are required for eye-specific patterning during retinogeniculate mapping. *Nat. Neurosci.*, *8*, 1022–1027.

Pfeiffenberger, C., Yamada, J., & Feldheim, D. A. (2006). Ephrin-As and patterned retinal activity act together in the development of topographic maps in the primary visual system. *J. Neurosci.*, *26*, 12873–12884.

Rakic, P. (1976). Prenatal genesis of connections subserving ocular dominance in the rhesus monkey. *Nature*, *261*, 467–471.

Rakic, P. (1981). Development of visual centres in the primate brain depends on binocular competition before birth. *Science*, *214*, 928–931.

Sanes, J. R., & Yamagata, M. (1999). Formation of lamina-specific synaptic connections. *Curr. Opin. Neurobiol.*, *9*, 79–87.

Sengpiel, F., & Kind, P. C. (2002). The role of activity in development of the visual system. *Curr. Biol.*, *12*, R818–826.

Shatz, C. J. (1983). The prenatal development of the cats' retinogeniculate pathway. *J. Neurosci.*, *3*, 482–499.

Shatz, C. J., & Stryker, M. P. (1978). Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *J. Physiol. (Lond.)*, *281*, 267–283.

Shatz, C. J., & Stryker, M. P. (1988). Prenatal tetrodotoxin infusion blocks segregation of retinogeniculate afferents. *Science*, *242*, 87–89.

Shook, B. L., & Chalupa, L. M. (1986). Organization of geniculocortical connections following prenatal interruption of binocular interactions. *Dev. Brain Res.*, *393*, 47–62.

Skaliora, I., Robinson, D. W., Scobey, R. P., & Chalupa, L. M. (1995). Properties of K+ conductances in cat retinal ganglion cells during the period of activity-mediated refinements in retinofugal pathways. *Eur. J. Neurosci.*, *7*, 1558–1568.

Skaliora, I., Scobey, R. P., & Chalupa, L. M. (1993). Prenatal development of excitability in cat retinal ganglion cells: Action potentials and sodium currents. *J. Neurosci.*, *13*, 313–323.

Stellwagen, D., & Shatz, C. J. (2002). An instructive role for retinal waves in the development of retinogeniculate connectivity. *Neuron*, *33*, 357–367.

Stevens, B., Allen, N. J., Vasquez, L. E., Christopherson, K. S., Nouri, N., Micheva, K. D., et al. (2007). The classical complement cascade mediates developmental synapse elimination. *Cell*, *131*, 1164–1178.

Stryker, M. P., & Harris, W. A. (1986). Binocular impulse blockade prevents the formation of ocular dominance columns in cat visual cortex. *J. Neurosci.*, *6*, 2117–2133.

Torborg, C. L., Hansen, K. A., & Feller, M. B. (2005). High frequency, synchronized bursting drives eye-specific segregation of retinogeniculate projections. *Nat. Neurosci.*, *8*, 72–78.

Wang, G. Y,, Ratto, G., Bisti, S., & Chalupa, L. M. (1997). Functional development of intrinsic properties in ganglion cells of the mammalian retina. *J. Neurophysiol.*, *78*, 2895–2903.

Warland, D. K., Huberman, A. D., & Chalupa, L. M., (2006). Dynamics of spontaneous activity in the fetal macaque retina during development of retinogeniculate pathways. *J. Neurosci.*, *26*, 5190–5197.

Weliky, M., & Katz, L. C. (1999). Correlational structure of spontaneous neuronal activity in the developing lateral geniculate nucleus in vivo. *Science*, *285*, 599–604.

Wiesel, T. N., & Hubel, D. H. (1965a). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J. Neurophysiol.*, *28*, 1029–1040.

Wiesel, T. N., & Hubel, D. H. (1965b). Extent of recovery from the effects of visual deprivation in kittens. *J. Neurophysiol.*, *28*, 1060–1072.

Wong, R. O. (1999). Retinal waves and visual system development. *Annu. Rev. Neurosci.*, *22*, 29–47.

Wong, R. O., Meister, M., & Shatz, C. J. (1993). Transient period of correlated bursting activity during development of the mammalian retina. *Neuron*, *11*, 923–938.

Wong, R. O., & Oakley, D. M. (1996). Changing patterns of spontaneous bursting activity of on and off retinal ganglion cells during development. *Neuron*, *16*, 1087–1095.

Zhang, L. I., & Poo, M. M. (2001). Electrical activity and development of neural circuits. *Nat. Neurosci.*, Suppl., 1207–1214.

Zhou, Z. J., & Zhou, D. (2000). Coordinated transitions in neurotransmitter systems for the initiation and propagation of spontaneous retinal waves. *J. Neurosci.*, *20*, 6570–6577.

5 Brain Changes Underlying the Development of Cognitive Control and Reasoning

SILVIA A. BUNGE, ALLYSON P. MACKEY, AND KIRSTIE J. WHITAKER

ABSTRACT What precisely is changing over time in a child's brain leading to improved control over his or her thoughts and behavior? This chapter investigates neural mechanisms that develop through childhood and adolescence and underlie changes in working memory, cognitive control, and reasoning. The effects of age and experience on specific cognitive functions are discussed with respect to functional brain imaging studies, highlighting the importance of interactions between prefrontal and parietal cortices in cognitive control and high-level cognition.

What precisely is changing over time in a child's brain, leading to improved control over his or her thoughts and behavior? Throughout childhood and adolescence, we improve at organizing our thoughts, working toward long-term goals, ignoring irrelevant information that could distract us from these goals, and controlling our impulses—in other words, we exhibit improvements in *executive function* or *cognitive control* (Diamond, 2002; Zelazo, Craik, & Booth, 2004; Casey, Tottenham, Liston, & Durston, 2005). By the same token, we exhibit increased facility over this age range in tackling novel problems and reasoning about the world—a capacity referred to as fluid reasoning (Cattell & Bernard, 1971). Both the capacity to consciously control our thoughts and actions and the capacity to reason effectively rely on *working memory*, or the ability to keep relevant information in mind as needed to carry out an immediate goal.

Neuroscientific research is being conducted to better understand the changes in brain structure and function that underlie improved cognitive control and fluid reasoning during child and adolescent development. More specifically, researchers seek to determine how the neural mechanisms underlying specific cognitive functions change with age, how they differ among individuals, and how they are affected by experience.

SILVIA A. BUNGE Helen Wills Neuroscience Institute and Department of Psychology, University of California at Berkeley, Berkeley, California

ALLYSON P. MACKEY and KIRSTIE J. WHITAKER Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, California

We begin this chapter with a brief summary of changes in brain structure, focusing primarily on prefrontal and parietal cortices, the brain regions that have been most closely associated with goal-directed behavior. We then provide an overview of functional brain imaging studies focusing on age-related changes in working memory, cognitive control, and fluid reasoning over childhood and adolescence. Because working memory and cognitive control development have been discussed extensively elsewhere (Munakata, Casey, & Diamond, 2004; Rubia & Smith, 2004; Casey et al., 2005; Bunge & Wright, 2007), a relatively greater emphasis is placed on recent studies focusing on the development of fluid reasoning.

Structural brain development

The brain undergoes major structural and functional changes over childhood and adolescence that may, in part, explain changes in behavior and cognition. As explained in chapter 2, by Kostović and Judaš, rapid changes occur at the neuronal level in the prefrontal cortex (PFC) in the first few years of life, followed by slower, protracted changes through adolescence (Petanjek, Judas, Kostovic, & Uylings, 2008). While brain changes at the cellular level can be examined only in postmortem brain tissue, advances in neuroimaging techniques have made it possible to study gross anatomical development in vivo. Structural magnetic resonance imaging (MRI) methods make it possible to quantify age-related changes in cortical thickness (Sowell et al., 2007), in the volume of specific brain structures (Gogtay et al., 2006), and in the thickness and coherence of white matter tracts connecting distant brain regions to one another (Giedd et al., 1999; Klingberg, Vaidya, Gabrieli, Moseley, & Hedehus, 1999).

Cortical thickness follows an inverted U-shaped pattern over development. Up to middle childhood (ages 8 to 12), increased thickness of the gray matter at the surface of the brain reflects increased density of neurons and dendrites. Thereafter, decreased gray matter thickness reflects the pruning of excess dendrites and neurons, as well as increased

myelination of axonal projections to these neurons (Giedd, 2004).

The developmental trajectory of changes in cortical thickness varies across brain structures. In PFC and parietal cortex, gray matter volume peaks around age 10–12 (Giedd, 2004). Thereafter, gray matter loss occurs at different rates in different subregions of the PFC, and it is considered one index of the time course of maturation of a region (Sowell et al., 2003). Within the PFC, gray matter reduction is completed earliest in the orbitofrontal cortex, followed by the ventrolateral PFC (VLPFC) and then by the dorsolateral PFC (DLPFC) and rostrolateral PFC (RLPFC) (Gogtay et al., 2004; O'Donnell, Noseworthy, Levine, & Dennis, 2005). Differences in maturational time course between prefrontal subregions could help account for differences in the rate of development of distinct cognitive control processes (Bunge & Zelazo, 2006; Crone et al., 2006).

Developmental changes in interregional connectivity have been studied with an MRI-based method known as diffusion tensor imaging (DTI). Research using DTI has shown that strengthening of frontal-parietal networks is associated with improved performance on working memory tasks (Olesen, Nagy, Westerberg, & Klingberg, 2003; Nagy, Westerberg, & Klingberg, 2004). An age-related increase in frontostriatal tract coherence has also been associated with more efficient recruitment of cognitive control (Liston et al., 2006).

In summary, both cortical pruning within prefrontal and parietal regions and increased neuronal connectivity within and between these and other regions are likely to underlie improvements in cognitive control and fluid reasoning during development. The relationships between behavioral improvements and changes in brain structure and brain function have been explored in recent studies of working memory, as described in the next section.

Working memory development

Working memory is the brain's "mental blackboard," allowing information—either sensory inputs or memories—to be held in mind and manipulated (Miller, Galanter, & Pribram, 1960; Goldman-Rakic, 1992). Considered a central component of human cognition, the maturation of working memory is critical for the development of language comprehension, mental calculation, cognitive control, and fluid reasoning. Although children as young as 5 years do not differ from adults in sensorimotor tasks, performance on tasks that rely on the retention and manipulation of information, such as spatial memory span and Tower of London, improves over childhood (Luciana & Nelson, 1998). Children's performance is critically moderated by task difficulty: their accuracy rapidly declines as the demands of the task become more rigorous and they make more errors. As working memory improves, children likewise improve on tests of cognitive control and fluid reasoning.

Working memory for different types of information is mediated by interactions between domain-specific brain regions and regions in PFC and parietal cortex (D'Esposito, 2007). It is the integration and refinement of these working memory circuits that underlies age improvements in the maintenance and manipulation of mental representations. In the next subsection, we provide a brief overview of fMRI studies examining age-related changes in working memory.

Visuospatial Working Memory Most fMRI studies on working memory development have focused on the ability to keep in mind a series of spatial locations (Casey et al., 1995; Thomas et al., 1999; Klingberg, Forssberg, & Westerberg, 2002a; Kwon, Reiss, & Menon, 2002; Scherf, Sweeney, & Luna, 2006). The superior frontal sulcus (SFS) and the intraparietal sulcus (IPS), which have been strongly implicated in adult visuospatial working memory, are increasingly engaged throughout childhood (Klingberg et al., 2002a; Kwon et al.). Across children, the level of fractional anisotropy in the frontoparietal white matter surrounding the SFS and IPS is positively correlated with visuospatial working memory scores (Nagy et al., 2004). Further, the coherence of these white matter tracts in the left hemisphere is greater among children (age 8–18 years) who exhibit the greatest activation in these regions (Olesen et al., 2003). Thus the brain network underlying effective visuospatial working memory is strengthened over development.

At a microscopic level, the increased engagement of SFS and IPS during a blood-oxygenation-level dependent (BOLD) fMRI visuospatial working memory task could be dependent on one or more cellular changes: neuronal pruning, increased myelination, and/or the strengthening of synaptic connections within or between brain regions. Klingberg and colleagues used computational methods to determine which of these processes are likely to support the development of visuospatial working memory (Edin, Macoveanu, Olesen, Tegner, & Klingberg, 2007). Their computational model of BOLD activation found that strengthened synaptic connectivity within and between brain regions was the most likely candidate for increase in activation in these regions between childhood and adulthood.

Just as core working memory networks are strengthened over childhood and adolescence, supporting networks that are not used by adults for working memory are weakened over this age range. Using a spatial working memory paradigm involving saccadic eye movements, Luna and colleagues showed that increased recruitment of core regions in DLPFC in the left hemisphere and parietal regions was accompanied by a weakening and eventual dismissal of a childhood compensatory circuit involving ventromedial PFC (Scherf et al., 2006). A qualitative shift was observed

in comparing children aged 10–13 and adolescents aged 14–17. Comparing adolescents to adults, the changes were more quantitative, evincing refinement of the visuospatial working memory network. This movement away from the childhood circuit to the more mature adult network is a common theme in developmental cognitive neuroscience, and it is further discussed in this chapter's section on cognitive control development.

Nonspatial Working Memory In addition to the fMRI studies of visuospatial working memory development, we would like to highlight a study on the development of nonspatial working memory, in which children aged 8–12, adolescents aged 13–17, and young adults were asked to remember a series of three nameable objects (figure 5.1*A*; Crone, Wendelken, Donohue, van Leijenhorst, & Bunge, 2006) . We consider first the pure maintenance condition of this study, in which participants were asked to verbally rehearse the items in the order in which they were presented. All three groups engaged highly overlapping sets of brain regions, indicating that the core working memory network was already in place by middle childhood. However, there was a positive correlation across participants between task accuracy and level of activation in left ventrolateral PFC (VLPFC), bilateral DLPFC, and bilateral superior parietal cortex.

Manipulation of Information in Working Memory This study of nonspatial working memory (Crone et al., 2006) also included a manipulation condition, in which participants were asked to reverse the order of the items in their head. Children aged 8–12 were disproportionately impaired relative to adolescents and adults on this manipulation condition relative to the pure maintenance condition. Further, children failed to engage right DLPFC and bilateral superior parietal cortex, regions linked with working memory manipulation, for this purpose. A qualitative shift in the circuitry underlying manipulation was observed from middle childhood onward, such that adolescents and

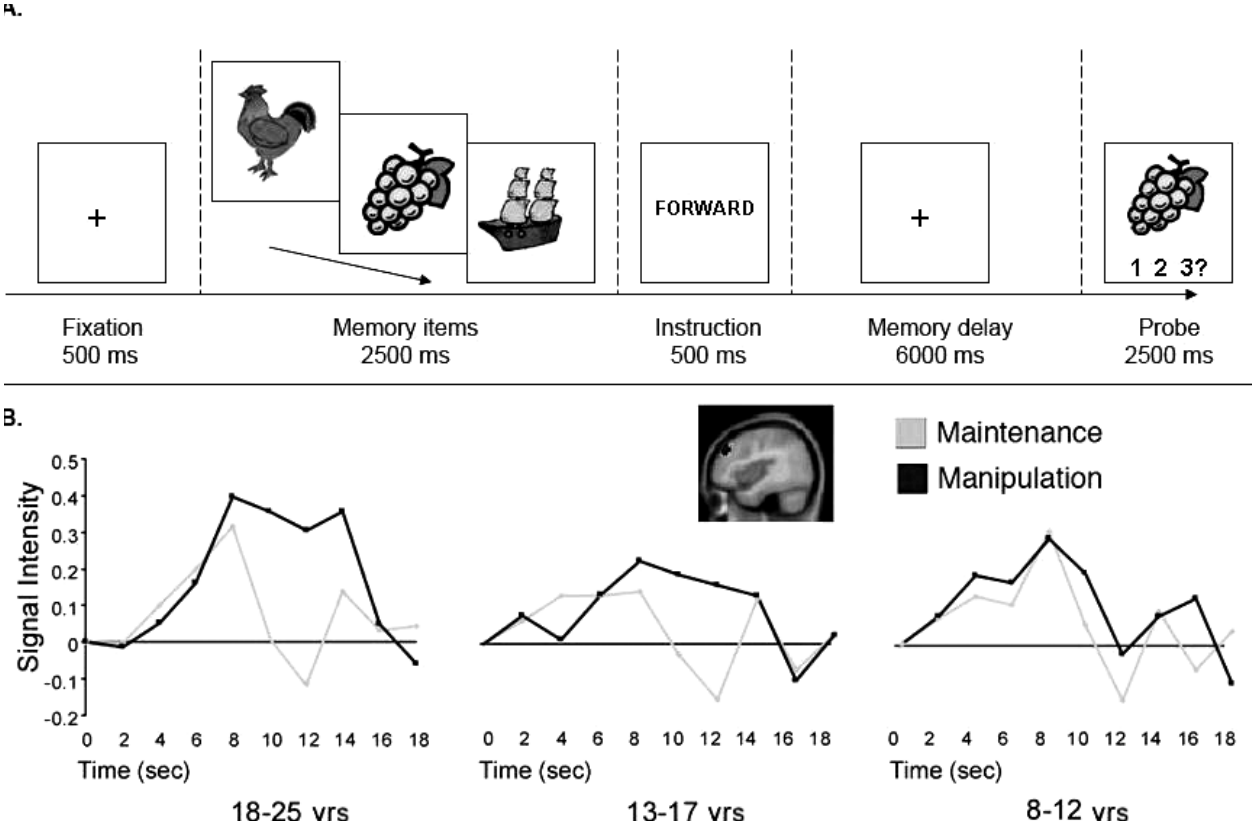


Figure 5.1 Development of nonspatial working memory and working memory manipulation. (*A*) Subjects were asked to remember three nameable objects, presented for 750 ms each and separated by a 250-ms fixation cross. After the last object the instruction "forward" or "backward" directed the participant to either mentally rehearse or reorder these objects during the 6,000-ms delay. Finally a probe object was presented and participants indicated with a button press whether it was first, second, or third object in the memorized sequence. Forward trials required pure maintenance, whereas backward trials required manipulation in addition to maintenance. (*B*) Group-averaged time courses for activation in the right DLPFC during the delay period show that adults and adolescents recruited this region more strongly during the harder manipulations trials, whereas children showed the same activation in DLPFC for both "forward" and "backward" tasks. (Reprinted with permission from Crone, Wendelken, Donohue, van Leijenhorst, & Bunge, 2006, copyright © 2006, National Academy of Sciences, USA.)

adults engaged an additional mechanism relative to children aged 8–12. Time-series correlational analyses showed that, for adults, right DLPFC was functionally correlated with bilateral parietal and premotor cortices during manipulation. In children, by contrast, right DLPFC activation during manipulation was correlated with regions that have not been associated previously with this function (unpublished analyses). Thus the brain network underlying manipulation in adults was not yet engaged by children aged 8–12. Importantly, it is not the case that children failed to engage these brain regions during task performance. Indeed, children engaged DLPFC and parietal cortex at encoding and retrieval of items in working memory; they simply failed to sufficiently engage the circuitry that supports manipulation at the time when it was required to reverse the order of objects in working memory (figure 5.1*B*).

These investigations have shown how the development of PFC and parietal cortex, as well as the connections between them, contributes to increased ability to maintain and manipulate information online. In turn, an increase in working memory capacity contributes to improvements in a variety of cognitive functions, including cognitive control and fluid reasoning.

Cognitive control development

One of the most obvious ways in which children mature behaviorally is that they become increasingly able to ignore irrelevant and distracting information and control their impulses while working toward specific goals. The terms *executive function* and *cognitive control* refer to mental processes associated with the control of thought and action. Thus far, developmental research on cognitive control has been concerned with conscious, deliberate forms of control. Putative cognitive control functions are listed in box 5.1.

Box 5.1 Cognitive Control Functions

1. Selectively attending to relevant information (*selective attention*) and ignoring distracting stimuli or thoughts (*interference suppression/resolution*)
2. Selecting between competing response tendencies (*response selection)* and inhibiting inappropriate response tendencies (*response inhibition*)
3. Using contextual information to identify currently relevant information and appropriate responses (*rule/task-set representation*)
4. Reorganizing information currently held in working memory (*manipulation, updating*)
5. Flexibly switching between tasks and performing two tasks concurrently (*task-switching, dual-task performance*)
6. Monitoring one's own actions and the consequences of these actions (*performance monitoring, error/feedback processing)*

Cognitive Control Development: Changes in One or More Neural Circuits? A key question in developmental research has been whether age-related changes in cognitive control are associated with the development of a single mechanism, such as the capacity to store or process information (Case, 1992; Dempster, 1993), or with a set of mechanisms (Welsh, Butters, Hughes, Mohs, & Heyman, 1991). Behavioral studies suggest that some of these abilities may mature at different rates. For example, the ability to inhibit a motoric response matures earlier than the ability to inhibit a response when the task additionally requires selective attention (van den Wildenberg & van der Molen, 2004). Likewise, the ability to switch between task rules develops earlier than the ability to keep a difficult rule online (Crone, Wendelken, Donohue, Honomichi, & Bunge, 2004). Recent advances using structural equation modeling indicate that working memory, task switching, and response inhibition are separable latent constructs with distinct developmental trajectories (Brocki & Bohlin, 2004; Huizinga, Dolan, & van der Molen, 2006). Thus behavioral research provides hints that different cognitive control functions may have separable neurodevelopmental trajectories.

In a recent study of cognitive network development, Fair, Dosenbach, and colleagues (2007) show that, initially, the strongest connections between frontoparietal gray matter are anatomically close together. As these regions mature, however, the connections become more functionally relevant and reach further afield, presumably to engage the most effective network for cognitive control (figure 5.2).

The protracted myelination of the white matter tracts connecting the regions necessary for cognitive control over childhood and adolescence (Spear, 2007) may explain the compensatory network required for children to complete these tasks. If long-range projections are not sufficiently insulated by the myelin sheath, they will be unable to communicate with the functionally relevant networks utilized by the adult brain. Eventually, during adolescence, the myelination is sufficient to allow a transition from the local, compensatory mechanism to a more diffuse, adult, effective system.

In the following sections, we highlight a few of the many brain-imaging studies that have examined neurodevelopmental changes in cognitive control.

Response selection and inhibition As noted previously, improvements in working memory manipulation—a form of cognitive control—are associated with an increase in lateral PFC activation with age (Crone et al., 2004). For some cognitive control tasks, however, maturation is instead associated with a decrease in lateral PFC recruitment. For example, a large study involving participants between 8 and 27 years of age by Luna and colleagues (Velanova, Wheeler, & Luna, 2008) examined functional maturational changes associated with

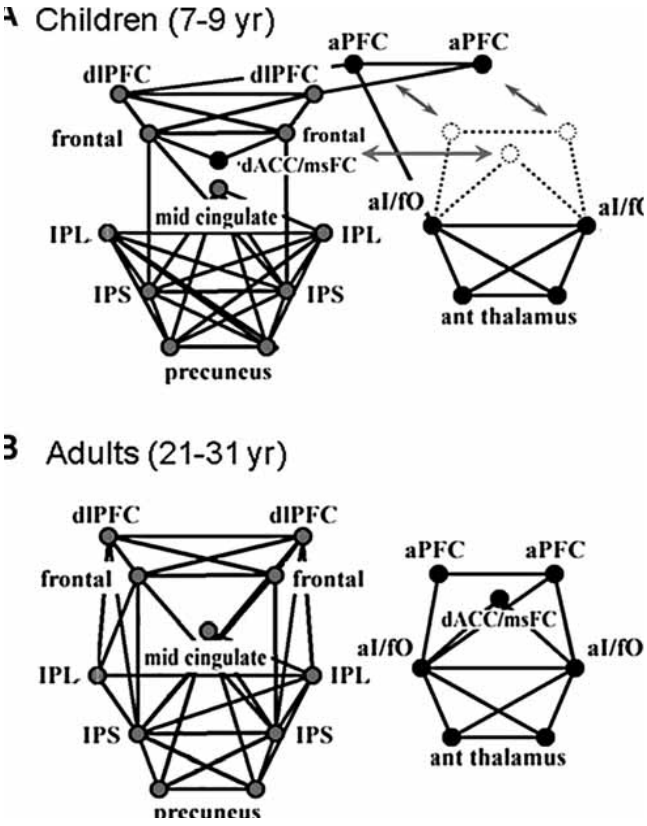


FIGURE 5.2 Development of distinct cognitive control networks through childhood and adolescence. Regions previously identified as pertaining to putative task control were analyzed for pairwise temporal BOLD correlations in (*A*) children and (*B*) adults. Right-side ROIs are displayed on the right of each graph and anterior ROIs at the top of each graph. Whereas adults demonstrate two separate control networks, children show a connection between them. Their networks are connected by a bridge between the anterior PFC and DLPFC, and the dACC and medial superior frontal cortex were incorporated into the frontoparietal network. In addition, children lacked connections between the DLPFC and the IPS and inferior parietal lobule. The two separate networks seen in adults are proposed to interpret cues, implement top-down control, and process bottom-up feedback, but use different mechanisms and over different temporal scales. (Reprinted with permission from Fair et al., 2007, copyright © 2007 National Academy of Sciences, USA.)

performance on an antisaccade task. In this response inhibition task, participants must move their eyes away from a visual stimulus that appears suddenly on the screen, resisting the urge to look toward it. The researchers observed an age-related shift away from reliance on DLPFC, toward posterior parietal regions (figure 5.2). Consistent with other work, this finding indicates a shift away from childhood compensatory mechanisms toward the more effective adult networks.

In addition to this antisaccade study, a number of other brain-imaging studies have focused on age-related improvements in the ability to select between competing response choices and inhibit inappropriate response tendencies (see Munakata, Casey, & Diamond, 2004, and Bunge & Wright, 2007, for reviews). These studies have made use of a variety of paradigms, including the well-known Stroop (Adleman, et al., 2002; Schroeter, Zysset, Wahl, & von Cramon, 2004; Marsh et al., 2006), go-no-go (Rubia, Smith, Taylor, & Brammer, 2007; Rubia et al., 2001; Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002; Durston et al., 2002; Tamm, Menon, & Reiss, 2002; Booth et al., 2003; Lamm, Zelazo, & Lewis, 2006), and flanker tasks (Bunge et al.; Lamm et al.; Rubia et al., 2006). In task-switching paradigms, the currently relevant task rule changes without warning, and it is necessary to suppress the response to the previous rule and also to retrieve the new rule from memory (Crone et al., 2004). These studies indicate that overlapping but distinct circuits involving regions of PFC, parietal cortex, and basal ganglia are involved in various cognitive control tasks (Rubia, et al., 2001).

Performance monitoring Luna and colleagues used the antisaccade task to examine not only the development of response inhibition, but also the development of performance monitoring (Velanova et al., 2008). The dorsal anterior cingulate cortex (dACC), known to play a central role in performance monitoring (Ford, Goltz, Brown, & Everling, 2005; Polli et al., 2005), was more strongly engaged on error trials in adults than in children or adolescents (figure 5.3). Increased performance monitoring, supported by dACC, is likely to contribute to the observed improvements in inhibitory cognitive control with age.

In summary, cognitive control is considered to comprise a variety of different putative processes, and a few studies have made attempts to compare the developmental time course of specific control processes (e.g., Bunge et al., 2002; Crone et al., 2006; Rubia et al., 2006). However, it remains to be seen if these different behavioral abilities are in fact different underlying neural substrates and if they develop over separable trajectories. Further work must be undertaken to determine the relative independence and interactions of the many cognitive control capabilities throughout development.

Fluid reasoning development

Fluid reasoning is the capacity to think logically and solve problems in novel situations (Cattell & Bernard, 1971). The concept of fluid reasoning is integral to theories of human intelligence (Horn & Cattell, 1967; Cattell, 1987; Horn, 1988; Carroll, 1997; McArdle & Woodcock 1998; Gray, Chabris, & Braver, 2003). Compared to crystallized, or knowledge-based, abilities it is thought to have a stronger neurobiological and genetic component, leading to the belief that it is less dependent on experience. However, some

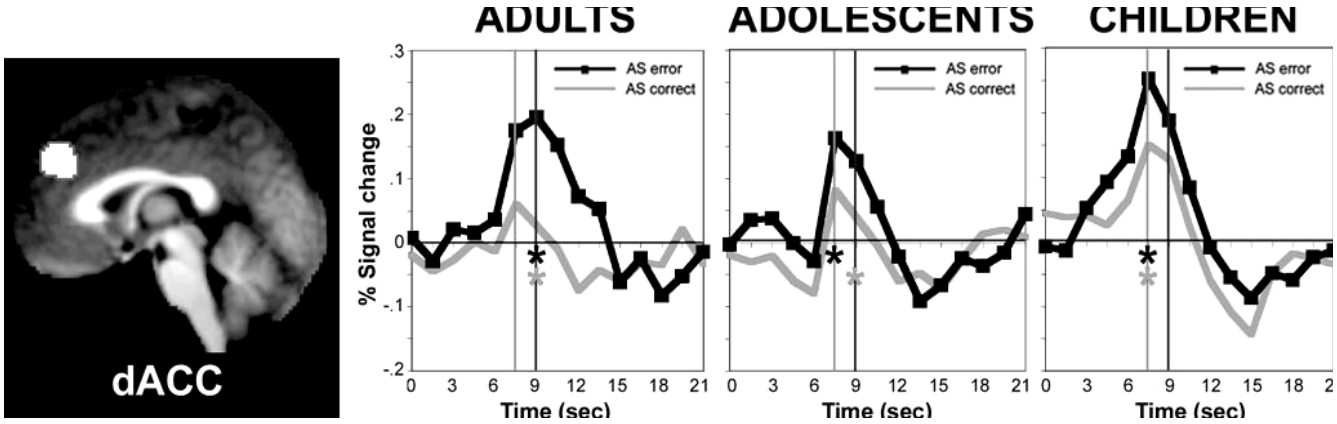


FIGURE 5.3 Development of response inhibition and performance monitoring. Velanova, Wheeler, and colleagues (2008) demonstrated that dACC showed significantly greater modulation during error versus correct trials in an antisaccade task. The task required subjects to inhibit the prepotent tendency to look toward the stimulus and look in the opposite direction. The time course of activation within the dACC, illustrated here, not only shows the difference between activity during error and correct trials but also that adults exhibit greater differential activity than adolescents and children. For each age group, black asterisks mark the time point showing mean maximal peak activity for error trials, and gray asterisks mark the time point showing mean maximal differences between error and correct trials. (Reprinted with permission from Velanova, Wheeler, & Luna, 2008, copyright Oxford University Press.)

evidence suggests that it is indeed sensitive to cultural and environmental influences (Flynn, 2007).

The development of reasoning ability is central to understanding cognitive development as a whole, because it serves as scaffolding for many other cognitive functions (Cattell, 1987; Blair, 2006). Fluid reasoning has been identified as a leading indicator of changes in crystallized abilities (McArdle, 2001). It strongly predicts changes in quantitative ability (Ferrer & McArdle, 2004) and reading (Ferrer et al., 2007) among children aged 5 to 10. Fluid reasoning ability even predicts performance through college and in cognitively demanding occupations (Gottfredson, 1997).

One form of fluid reasoning is relational reasoning: the ability to consider relationships between multiple distinct mental representations (Gentner, 1983; Hummel & Holyoak, 1997). Analogical reasoning, more specifically, involves abstracting a relationship between familiar items and applying it to novel representations (Gentner, 1988; Goswami, 1989). In other words, forming analogies allows us to determine general principles from specific examples and to establish connections between previously unrelated pieces of information. Analogical thought is an important means by which cognition develops (Goswami, 1989; R. Brown & Marsden, 1990). For example, children use analogies to learn new words and concepts by association with previously learned information (Gentner, 1983).

WHEN DOES REASONING ABILITY DEVELOP? Historically, theories of reasoning development focused on children's limitations. Piaget claimed that, before the stage of formal operations around age 11, children are not capable of mentally representing the relations necessary to solve analogies (Inhelder & Piaget, 1958). When Piaget and his colleagues showed children pictorial problems of the form "A is to B as C is to . . . ?" and asked them to find the D term among a set of pictures, he found that children often chose items that were perceptually or semantically related to the C item (Piaget, Montangero, & Billeter, 1977). Sternberg and colleagues found similar limitations in young children's analogical reasoning, observing an overreliance on lower-order relations during analogical problem solving (Sternberg & Nigro, 1980; Sternberg & Downing, 1982). It has been argued that children as young as 3 years old can solve simple analogies as long as they are familiar with the objects involved and understand the relevant relations (Goswami, 1989), but improvements in analogical reasoning are observed throughout childhood and adolescence (Sternberg & Rifkin, 1979; Richland, Morrison, & Holyoak, 2006).

Fluid reasoning ability seems to be a distinct cognitive function, rising and falling at its own rate across the life span (Cattell, 1987). It follows a different developmental trajectory than crystallized abilities, supporting the idea that these are separable cognitive functions (Horn, 1991; Schaie, 1996; McGrew, 1997). Fluid reasoning capacity increases very rapidly until late adolescence and early adulthood, peaking at around age 22 and declining thereafter (McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002).

ASSESSING REASONING ABILITY One of the most commonly used measures of fluid reasoning ability is the Raven's Progressive Matrices test (RPM), a classic visuospatial task that can be administered to both children and adults (Raven, 1941). This test is considered an excellent measure of fluid reasoning ability (Kline, 1993) and of intellectual ability overall (Wechsler & Stone, 1945).

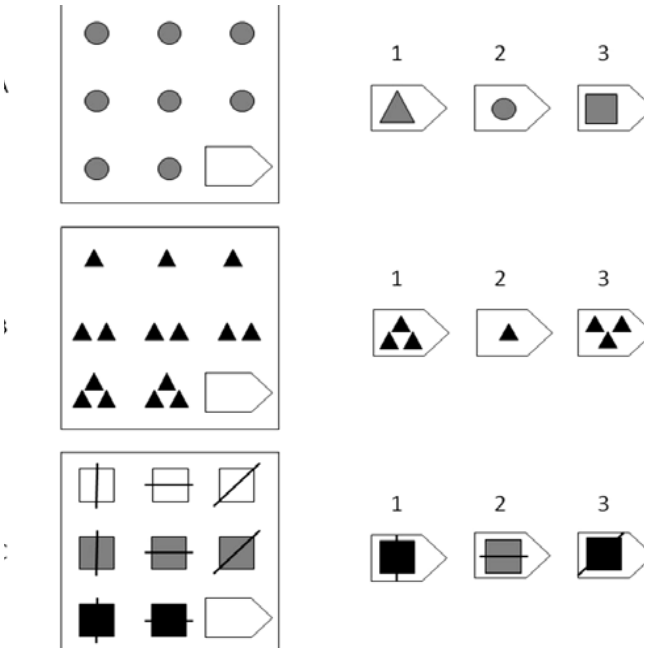


FIGURE 5.4 Sample problem similar to Raven's Progressive Matrices. (*A*) Zero-relational problem (REL-0) that requires only perceptual matching (Answer: 2). (*B*) One-relational problem (REL-1) that involves consideration of change in either the vertical or horizontal direction (Answer: 1). (*C*) Two-relational problem (REL-2) that requires attention to change in both the vertical and horizontal directions (Answer: 3).

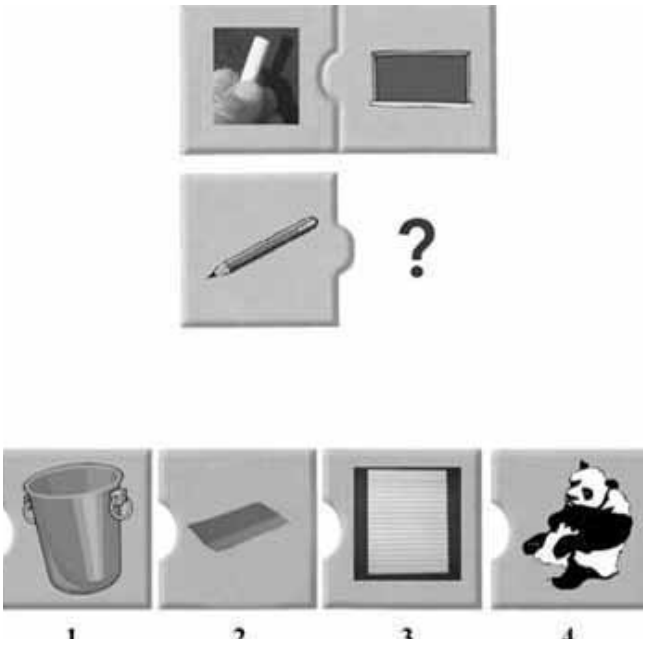


FIGURE 5.5 Sample visual analogy problem. On this type of problem, subjects must consider the relationship between the top two images and choose the image that completes the bottom analogy (Answer: 3; 2 is the semantic lure).

As illustrated in figure 5.4, the RPM includes zero-relational (REL-0), one-relational (REL-1), and two-relational (REL-2) problems. REL-0 problems require only perceptual matching. REL-1 problems require subjects to consider either vertical or horizontal changes (or spatial relations) across figures in a 3×3 grid to infer the missing piece in the bottom right corner. REL-2 problems require subjects to process changes in both the horizontal and vertical directions simultaneously in order to choose the missing piece. These problems are the most difficult because they require the integration of two visuospatial relations.

Analogical reasoning can be assessed behaviorally and in an MRI scanner with propositional analogy problems involving either words or pictures of nameable objects. The visual analogy task used in a recent fMRI study from our laboratory (Wright, Matlen, Baym, Ferrer, & Bunge, 2008) requires children to select which of four pictures completes an analogy. The answer choices for these problems include perceptual and semantic lures (figure 5.5).

NEURAL CORRELATES OF FLUID REASONING Brain regions important for fluid reasoning have been identified through studies of patients with impaired reasoning ability and neuroimaging studies of healthy adults. Research on patients in the early stages of frontotemporal dementia (FTD) has shown that reasoning is differentially affected based on the brain areas most compromised by the disease. Patients with frontal-variant FTD make errors on analogical reasoning problems related to limited working memory and trouble inhibiting inappropriate responses. In contrast, patients with temporal-variant FTD are profoundly impaired on analogical reasoning problems as a result of semantic memory loss (Morrison et al., 2004). Another study of patients with prefrontal damage revealed that these patients have a specific deficit in relational integration as compared to patients with anterior temporal lobe damage, who are more impaired on tests of episodic and semantic memory (Waltz et al., 1999).

Imaging research has narrowed down the region in PFC responsible for relational integration to the most anterior part of lateral PFC (RLPFC). Functional MRI studies of reasoning, including the RPM task (Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997; Christoff et al., 2001; Kroger et al., 2002) and a verbal propositional analogy task (Bunge, Wendelken, Badre, & Wagner, 2005), have implicated RLPFC in problems that require joint consideration of multiple relations. The other lateral PFC regions play roles in reasoning that are not specifically associated with relational complexity. DLPFC may support reasoning by organizing representations in working memory, selecting between competing response alternatives, and monitoring performance (Christoff et al.). Depending on the nature of the task, different brain regions contribute to fluid reasoning. Left VLPFC (Broca's area) supports reasoning by retrieving

semantic relations on propositional analogy problems (Bunge et al.; Wright et al., 2008). Likewise, hippocampus and parietal cortex may play a role in reasoning by representing individual relations in visuospatial tasks involving relational integration, including Raven's Progressive Matrices (Crone et al., 2009) and transitive inference problems (Wendelken & Bunge, under review).

Parietal cortex is consistently engaged in high-level cognitive tasks like the RPM (Gray et al., 2003) and shares strong connections with PFC (Petrides & Pandya, 1984; Fuster, 2002). A study of individual differences in reasoning ability in adults showed that stronger prefrontal and parietal recruitment on a difficult working memory task is associated with better fluid reasoning, as measured by an RPM-type task (Gray et al.). The level of activation in left lateral PFC and bilateral parietal cortex accounted for more than 99.9% of the relationship between fluid intelligence and working memory performance in these adults.

Taken together, the preceding studies suggest that maturation of RLPFC should lead to improvements in relational processing, while maturation of Broca's area, hippocampus, and parietal cortex should lead to better reasoning through improved representation of individual verbal and visuospatial relations.

How Does the Brain Change to Allow for Improvements in Reasoning Ability? The neuroimaging research described thus far has identified brain regions that contribute to reasoning in adults. However, researchers are just now beginning to track how these regions develop during childhood and how this development leads to improved fluid reasoning ability. As noted earlier in this chapter, in the section on structural brain development, DLPFC, RLPFC, and parietal cortex develop relatively slowly: cortical gray matter loss continues through the early twenties (Giedd, 2004). A study by Shaw and colleagues showed that the trajectory of changes in cortical thickness in several prefrontal regions differed across children with superior, high, and average intelligence (Shaw et al., 2006). Surprisingly, children with superior intelligence exhibited a delayed peak of cortical thickness in anterior PFC relative to the other groups, around age 11 as opposed to age 7–8 in children of average intelligence. The precise significance of this intriguing finding is as yet unclear. In particular, the role of environmental factors has not been explored; in this study sample, IQ was correlated with socioeconomic factors. However, this finding indicates that cognitive ability is related to the particular time course of cortical maturation in frontal regions, rather than the size of a given region at a specific age. This finding speaks to the unique insights that can be gained from longitudinal studies of brain development.

While structural imaging has provided critical insight into the neural changes that underlie reasoning development, functional neuroimaging is essential to understand how changes in brain function lead to changes in behavior. This section highlights the first two fMRI studies of reasoning ability in children.

Visual analogies In the first study, our laboratory (Wright et al., 2008) presented children aged 6–13 and young adults with visual analogy problems (figure 5.5). Children were capable of identifying analogous relationships between pairs of images, but made disproportionately more mistakes than adults on the analogy problems relative to 1-relational problems that required them to select from several images the one that was most semantically related to a cue image (figure 5.5*A*).

In left VLPFC, a region involved in the effortful retrieval of individual semantic relations between items (see, for example, Badre & Wagner, 2007), no consistent differences were observed between children and adults (Wright et al., 2008). However, older children did engage this region more strongly than younger children, indicating that left VLPFC contributes increasingly to controlled semantic retrieval between the ages of 6 and 13.

In bilateral RLPFC, the time-course analyses provided strong evidence for an immature activation profile in children (figure 5.6). The peak of activation in RLPFC occurred at least 4 seconds later for children than for adults, despite minimal differences in response times between the groups. In fact, for children, RLPFC activation peaked after the motor cortex activation associated with the behavioral response. Overall, consistent with a model whereby relatively more rostral PFC matures later than caudal PFC (Bunge & Zelazo, 2006), larger differences between children and adults were observed in RLPFC than in VLPFC. Changes in the function of RLPFC over childhood and adolescence may contribute to improvements in reasoning ability, and individual differences in RLPFC functioning may explain, at least in part, why some people have a greater capacity for fluid reasoning than others.

Raven's Progressive Matrices In the second study, our laboratory (Crone et al., 2009) presented children aged 8–12 and young adults with problems adapted from the Raven's Progressive Matrices (figure 5.4). Behaviorally, children made a disproportionate number of errors on the REL-2 problems relative to REL-1 problems, and their response times on the REL-2 problems did not differ from those of adults. This finding suggests that children selected responses for these difficult problems before adequately considering both dimensions of relational change.

In adults, RLPFC activation was greater for REL-2 problems than for REL-1 problems. While children also recruited RLPFC, they did not exhibit sustained preferential recruitment of RLPFC for REL-2 problems as compared with

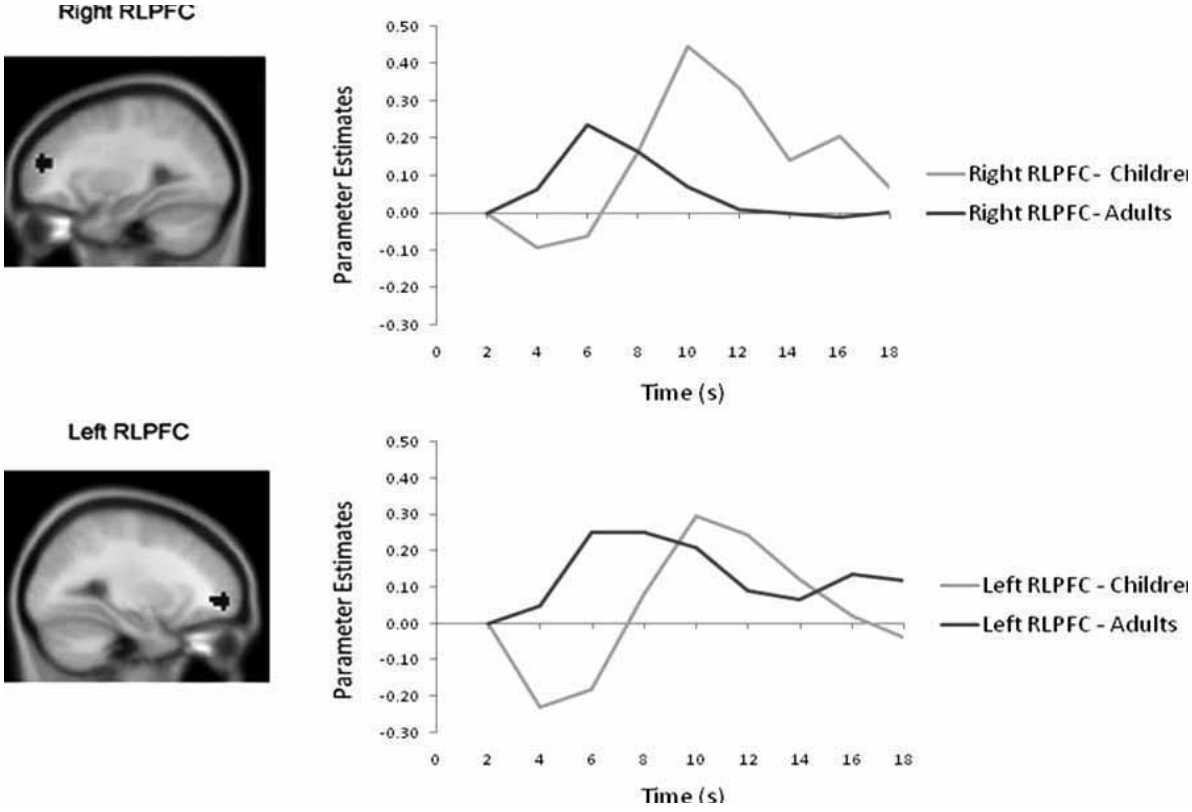


FIGURE 5.6 RLPFC regions of interest and time courses. On the left side of this image, the right and left RLPFC regions of interest are shown in a sagittal view. The right side displays time courses from these regions from baseline at 2 seconds through 18 seconds after trial onset for both children and adults. Bilaterally, the peak of activation occurs about 4 seconds later in children than in adults, and these regions even appear deactivated during the first few seconds of stimulus presentation.

REL-1 problems. Together with the response time data, this finding suggests that the children were more likely to treat the REL-2 problems similarly to REL-1 problems, considering only a single dimension of change. Activation of RLPFC associated with the REL-2 problems increases with age, indicating that development of RLPFC integration mechanism occurs, at least in part, over the age range (8–12 years) that was studied. Unlike RLPFC, the inferior parietal lobule was sensitive to the number of relations in adults and showed an immature pattern of activation in children. In summary, this study provides evidence that the development of reasoning is associated with functional changes in RLPFC in response to relational integration.

In summary, fluid reasoning ability comes online early in childhood but continues to develop through adolescence and even into adulthood. Intelligence in adults is related to connectivity between PFC and parietal cortex (Shaw et al., 2006). Structural neuroimaging studies (Giedd, 2004) have shown that development of these regions, PFC and parietal cortex, follows a prolonged developmental time course that matches behavioral data on reasoning in childhood (Richland et al., 2006). Initial functional neuroimaging studies have shown that children recruit brain regions similar to those that adults use to solve analogy problems, but with different patterns of activation suggesting functional immaturity (Wright et al., 2008; Crone et al., 2009).

Conclusions

A growing literature indicates that the increased recruitment of task-related regions in prefrontal and parietal regions contribute to improvements in goal-directed behavior over middle childhood and adolescence. The pattern of developmental changes in brain activation has been generally characterized as a shift from diffuse to focal activation (Durston, Davidson, et al., 2006) and from posterior to anterior activation (Rubia et al., 2007; T. Brown et al., 2005). Differences can be either quantitative, with one age group engaging a region more strongly or extensively than another, or qualitative, with a shift in reliance on one set of brain regions to another, or both (T. Brown et al., 2005; T. Brown, Petersen, & Schlagger, 2006; Rubia et al., 2007; Scherf et al., 2006; Badre & Wagner, 2007). Importantly, the precise pattern of change observed depends on the task, the ages being examined, and the brain region in question. By further characterizing neurodevelopmental changes in cognitive control processes within subjects and across a range of tasks, we hope to better understand the development of the human mind.

Current and Future Directions By around age 12, the ability to hold goal-relevant information in mind and use it to select appropriate actions is already adequate, although not fully mature. It is of great interest to track brain function associated with working memory and cognitive control earlier in childhood, when these abilities are first acquired. Optical imaging studies can be conducted from infancy onward, although the spatiotemporal resolution of this method is suboptimal. It is now possible to acquire fMRI data in children as young as four years of age (Cantlon et al., 2006), although not without challenges like head motion, low accuracy, and poor attention span.

An important future direction is to determine the extent to which observed age differences in brain activation reflect hard developmental constraints (e.g., the required anatomical network is simply not yet in place at a given age) as opposed to lack of experience with a given type of task or cognitive strategy. Training studies involving several age groups would allow us to investigate effects of age and effects of practice independently and to test whether inherent age differences in performance and brain activation are still present after substantial practice (Luna & Sweeney, 2004; Qin et al., 2004).

Thus far, all but one (Durston, Davidson, et al., 2006) of the published developmental fMRI studies on working memory or cognitive control have compared groups of individuals at different ages. While these cross-sectional studies are valuable, they provide only a coarse indicator of developmental change. It is also important to conduct longitudinal studies to characterize intra-individual changes in brain function with age.

To understand how goal-directed behavior is achieved, it will be necessary to know how PFC and parietal cortices interact with other brain regions. It is the maturation of a specific network, rather than a particular brain region, that determines how effectively a given brain process is carried out. Some information about these interactions can be gleaned from functional connectivity analyses of fMRI data. Another approach is to acquire fMRI and EEG data in the same group of participants, either in separate sessions or simultaneously (Debener et al., 2005). An important current and future direction for developmental neuroimaging studies is to examine developmental changes in interactions between brain regions, furthering the work of Fair and colleagues (2007) demonstrated in figure 5.2.

The examination of the normal developmental pathways of distinct control functions will be important for understanding sensitive periods in brain development. For example, damage to PFC in childhood has a much greater impact than does damage in adulthood, likely because this region is important for acquiring skills and knowledge during childhood (Eslinger, Flaherty-Craig, & Benton, 2004).

Finally, a better understanding of neurodevelopmental changes in healthy children will lead to insights into the reasons for impoverished goal-directed behavior in a number of neurodevelopmental disorders, such as attention-deficit/hyperactivity disorder (Vaidya et al., 2005; Durston, Mulder, et al., 2006) and Tourette syndrome (Peterson, Pine, Cohen, & Brook, 2001; Baym, Corbett, Wright, & Bunge, 2008).

REFERENCES

Adleman, N. E., Menon, V., Blasey, C. M., White, C. D., Warsofsky, I. S., Glover, G. H., & Reiss, A. L. (2002). A developmental fMRI study of the Stroop color-word task. *NeuroImage*, *16*(1), 61–75.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883–2901.

Baym, C. L., Corbett, B. A., Wright, S. B., & Bunge, S. A. (2008). Neural correlates of tic severity and cognitive control in children with Tourette syndrome. *Brain*, *131*(1), 165–179.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behav. Brain Sci.*, *29*(2), 109–125; discussion, 125–160.

Booth, J. R., Burman, D. D., Meyer, J. R., Lei, Z., Trommer, B. L., Davenport, N. D., Li, W., Parrish, T. B., Gitelman, D. R., & Mesulam, M. M. (2003). Neural development of selective attention and response inhibition. *NeuroImage*, *20*(2), 737–751.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Dev. Neuropsychol.*, *26*(2), 571–593.

Brown, R. G., & Marsden, C. D. (1990). Cognitive function in Parkinson's disease: From description to theory. *Trends Neurosci.*, *13*(1), 21–29.

Brown, T. T., Lugar, H. M., Coalson, R. S., Miezin, F. M., Petersen, S. E., & Schlagger, B. L. (2005). Developmental changes in human cerebral functional organization for word generation. *Cereb. Cortex*, *15*(3), 275–290.

Brown, T. T., Petersen, S. E., & Schlagger B. L. (2006). Does human functional brain organization shift from diffuse to focal with development? *Dev. Sci.*, *9*(1), 9–11.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. (2002). Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron*, *33*(2), 301–311.

Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cereb. Cortex*, *15*(3), 239–249.

Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Curr. Opin. Neurobiol.*, *17*(2), 243–250.

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Curr. Dir. Psychol. Sci.*, *15*(3), 118–121.

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biol*, *4*(5), e125.

Carroll, J. B. (1997). *The three-stratum theory of cognitive abilities*. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122–130). New York: Guilford.

Case, R. (1992). The role of the frontal lobes in the regulation of cognitive development. *Brain Cogn.*, *20*(1), 51–73.

Casey, B. J., Cohen, J. D., Jezzard, P., Turner, R., Noll, D. C., Trainor, R. J., Giedd, J., Kaysen, D., Hertz-Pannier, L., & Rapoport, J. L. (1995). Activation of prefrontal cortex in children during a nonspatial working memory task with functional MRI. *NeuroImage*, *2*, 221–229.

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends Cogn. Sci. 9*(3), 104–110.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam: North-Holland.

Cattell, R. B., & Bernard, R. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton Mifflin.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, *14*(5), 1136–1149.

Crone, E. A., Wendelken, C., Donohue, S., Honomichi, R., & Bunge, S. A. (2004). *Contributions of prefrontal subregions to developmental changes in rule use*. San Diego: Society for Neuroscience.

Crone, E. A., Wendelken, C., Donohue, S., van Leijenhorst, L., & Bunge, S. A. (2006). Neurocognitive development of the ability to manipulate information in working memory. *Proc. Natl. Acad. Sci. USA*, *103*(24), 9315–9320.

Crone, E. A., Wendelken, C., van Leijenhorst, L., Honomichl, R., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Dev. Sci*, *12*(1), 55–66.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K.. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *J. Neurosci.*, *25*(50), 11730–11737.

Dempster, F. N. (1993). Resistance to interference: Developmental changes in a basic processing mechanism. In M. L. Howe & H. R. Pasnak (Eds.), *Emerging themes in cognitive development:* Vol. 1. *Foundations* (pp. 3–27). New York: Springer-Verlag.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philos. Trans. R. Soc. London B Biol. Sci.*, *362*(1481), 761–772.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy and biochemistry. In S. A. Knight (Ed.), *Principles of frontal lobe function* (pp. 466–503). Oxford, UK: Oxford University Press.

Durston, S., Davidson, M. C., Tottenham, N., Galvan, A., Spicer, J., Fossella, J. A., & Casey, B. J. (2006). A shift from diffuse to focal cortical activity with development. *Dev. Sci.*, *9*(1), 1–8.

Durston, S., Mulder, M., Casey, B. J., Ziermans, T., & van Engeland, H. (2006). Activation in ventral prefrontal cortex is sensitive to genetic vulnerability for attention-deficit hyperactivity disorder. *Biol. Psychiatry*, *60*(10), 1062–1070.

Durston, S., Thomas, K. M., Yang, Y. H., Ulug, A. M., Zimmerman, R. D., & Casey, B. J. (2002). A neural basis for the development of inhibitory control. *Dev. Sci.*, *5*(4): F9–F16.

Edin, F., Macoveanu, J., Olesen, P., Tegner, J., & Klingberg, T. (2007). Stronger synaptic connectivity as a mechanism behind development of working memory–related brain activity during childhood. *J. Cogn. Neurosci.*, *19*(5), 750.

Eslinger, P. J., Flaherty-Craig, C. V., & Benton, A. L. (2004). Developmental outcomes after early prefrontal cortex damage. *Brain Cogn.*, *55*(1), 84–103.

Fair, D. A., Dosenbach, N. U., Church, J. A., Cohen, A. L., Brahmbhatt, S., Miezin, F. M., Barch, D. M., Raichle, M. E., Petersen, S. E., & Schlaggar, B. L. (2007). Development of distinct control networks through segregation and integration. *Proc. Natl. Acad. Sci. USA*, *104*(33), 13507–13512.

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Dev. Psychol.*, *40*(6), 935–952.

Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Dev. Psychol.*, *43*, 1460–1473.

Flynn, J. (2007). *What is intelligence?* New York: Cambridge University Press.

Ford, K. A., Goltz, H. C., Brown, M. R., & Everling, S. (2005). Neural processes associated with antisaccade task performance investigated with event-related fMRI. *J. Neurophysiol.*, *94*(1), 429–440.

Fuster, J. M. (2002). Frontal lobe and cognitive development. *J. Neurocytol.*, *31*(3–5), 373–385.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cogn. Sci.*, *7*, 155–170.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Dev.*, *59*, 47–59.

Giedd, J. N. (2004). Structural magnetic resonance imaging of the adolescent brain. *Ann. NY Acad. Sci.*, *1021*, 77–85.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nat. Neurosci.*, *2*(10), 861–863.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., 3rd, Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. USA*, *101*(21), 8174–8179.

Gogtay, N., Nugent, T. F., 3rd, Herman, D. H., Ordonez, A., Greenstein, D., Hayashi, K. M., Clasen, L., Toga, A. W., Giedd, J. N., Rapoport, J. L., & Thompson, P. M. (2006). Dynamic mapping of normal human hippocampal development. *Hippocampus*, *16*(8), 664–672.

Goldman-Rakic, P. S. (1992). Working memory and the mind. *Sci. Am.*, *267*(3), 110–117.

Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cogn. Dev.*, *4*, 251–268.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.

Gray, J. R., Chabris, C. F., & Braver, J. S. (2003). Neural mechanisms of general fluid intelligence. *Nat. Neurosci.*, *6*(3), 316–322.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 645–685). New York: Academic Press.

Horn, J. L. (1991). Measurement of intellectual capacities: A review in theory. In K. S. McGrew, J. K. Werder, & R. W.

Woodcock (Eds.), *Woodcock-Johnson technical manual* (pp. 197–246). Allen, TX: DLM Teaching Resources.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychol. (Amst.)*, *26*, 107–129.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychol. Rev.*, *10*(3), 427–466.

Inhelder, B. P., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Kline, P. (1993). *The handbook of psychological testing*. London: Routledge.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002a). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *J. Cogn. Neurosci.*, *14*(1), 1–10.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002b). Training of working memory in children with ADHD. *J. Clin. Exp. Neuropsychol.*, *24*(6), 781–791.

Klingberg, T., Vaidya, C. J., Gabrieli, J. D., Moseley, M. E., & Hedehus, M. (1999). Myelination and organization of the frontal white matter in children: A diffusion tensor MRI study. *NeuroReport*, *10*(13), 2817–2821.

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cereb. Cortex*, *12*(5), 477–485.

Kwon, H., Reiss, A. L., & Menon, V. (2002). Neural basis of protracted developmental changes in visuo-spatial working memory. *Proc. Natl. Acad. Sci. USA*, *99*(20), 13336–13341.

Lamm, C., Zelazo, P. D., & Lewis, M. D. (2006). Neural correlates of cognitive control in childhood and adolescence: Disentangling the contributions of age and executive function. *Neuropsychologia*, *44*(11), 2139.

Liston, C., Watts, R., Tottenham, N., Davidson, M. C., Niogi, S., Ulug, A. M., & Casey, B. J. (2006). Frontostriatal microstructure modulates efficient recruitment of cognitive control. *Cereb. Cortex*, *16*(4), 553–560.

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, *36*(3), 273–293.

Luna, B., & Sweeney, J. A. (2004). The emergence of collaborative brain function: FMRI studies of the development of response inhibition. *Ann. NY Acad. Sci.*, *1021*, 296–309.

Marsh, R., Zhu, H., Schultz, R. T., Quackenbush, G., Royal, J., Skudlarski, P., & Peterson, B. S. (2006). A developmental fMRI study of self-regulatory control. *Hum. Brain Mapping*, *27*(11), 848–863.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. DuToit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future: A Festschrift in honor of Karl Jöreskog* (pp. 341–380). Lincolnwood, IL: Scientific Software International.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analysis of growth and decline of multiple intellectual abilities over the lifespan. *Dev. Psychol.*, *38*(1), 113–142.

McArdle, J. J., & Woodcock, J. R. (1998). *Human cognitive abilities in theory and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In K. Pribram, D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 131–150). New York: Guilford.

Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *J. Cogn. Neurosci.*, *16*(2), 260–271.

Munakata, Y., Casey, B. J., & Diamond, A. (2004). Developmental cognitive neuroscience: Progress and potential. *Trends Cogn. Sci.*, *8*(3), 122–128.

Nagy, Z., Westerberg, H., & Klingberg, T. (2004). Maturation of white matter is associated with the development of cognitive functions during childhood. *J. Cogn. Neurosci.*, *16*(7), 1227–1233.

O'Donnell, S., Noseworthy, M. D., Levine, B., & Dennis, M. (2005). Cortical thickness of the frontopolar area in typically developing children and adolescents. *NeuroImage*, *24*(4), 948–954.

Olesen, P. J., Nagy, Z., Westerberg, H., & Klingberg, T. (2003). Combined analysis of DTI and fMRI data reveals a joint maturation of white and grey matter in a fronto-parietal network. *Brain Res. Cogn. Brain Res.*, *18*(1), 48–57.

Petanjek, Z., Judas, M., Kostovic, I., & Uylings, H. B. (2008). Lifespan alterations of basal dendritic trees of pyramidal neurons in the human prefrontal cortex: A layer-specific pattern. *Cereb. Cortex*, *18*(4), 915–929.

Peterson, B. S., Pine, D. S., Cohen, P., & Brook, J. S. (2001). Prospective, longitudinal study of tic, obsessive-compulsive, and attention-deficit/hyperactivity disorders in an epidemiological sample. *J. Am. Acad. Child Adolesc. Psychiatry*, *40*(6), 685–695.

Petrides, M., & Pandya, D. N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.*, *228*(1), 105–116.

Piaget, J. M., Montangero, J., & Billeter, J. (1977). La formation des correlates. In J. Piaget (Ed.), *Recherches sur l'abstraction réfléchissante* (pp. 115–129). Paris: Presses Universitaires de France.

Polli, F. E., Barton, J. J., Cain, M. S., Thakkar, K. N., Rauch, S. L., & Manoach, D. S. (2005). Rostral and dorsal anterior cingulate cortex make dissociable contributions during antisaccade error commission. *Proc. Natl. Acad. Sci. USA*, *102*(43), 15700–15705.

Prabhakaran, V., Smith, J. A., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cogn. Psychol.*, *33*, 43–63.

Qin, Y., Carter, C. S., Silk, E. M., Stenger, V. A., Fissell, K., Goode, A., & Anderson, J. R. (2004). The change of the brain activation patterns as children learn algebra equation solving. *Proc. Natl. Acad. Sci. USA*, *101*(15), 5686–5691.

Raven, J. C. (1941). Standardization of progressive matrices. *Br. J. Med. Psychol.*, *19*, 137–150.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *J. Exp. Child Psychol.*, *94*(3), 249–273.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., Simmons, A., Williams, S. C. R., Giampietro, V., Andrew, C. M., & Taylor, E. (2001). Mapping

motor inhibition: Conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage*, *13*(2), 250–261.

Rubia, K., & Smith, A. (2004). The neural correlates of cognitive time management: A review. *Acta Neurobiol. Exp.*, (*Warsz.*) *64*, 329–340.

Rubia, K., Smith, A. B., Taylor, E., & Brammer, M. (2007). Linear age-correlated functional development of right inferior fronto-striato-cerebellar networks during response inhibition and anterior cingulate during error-related processes. *Hum. Brain Mapping*, *28*, 1163–1177.

Rubia, K., Smith, A. B., Wooley, J., Nosarti, C., Heyman, I., Taylor, E., & Brammer, M. (2006). Progressive increase of frontostriatal brain activation from childhood to adulthood during event-related tasks of cognitive control. *Hum. Brain Mapping*, *27*(12), 973–993.

Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study*. New York: Cambridge University Press.

Scherf, K. S., Sweeney, J. A., & Luna, B. (2006). Brain basis of developmental change in visuospatial working memory. *J. Cogn. Neurosci.*, *18*(7), 1045.

Schroeter, M. L., Zysset, S., Wahl, M., & von Cramon, D. Y. (2004). Prefrontal activation due to Stroop interference increases during development—An event-related fNIRS study. *NeuroImage*, *23*(4), 1317–1325.

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., & Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature*, *440*(7084), 676–679.

Sowell, E. R., Peterson, B. S., Kan, E., Woods, R. P., Yoshii, J., Bansal, R., Xu, D., Zhu, H., Thompson, P. M., & Toga, A. W. (2007). Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. *Cereb.Cortex*, *17*(7), 1550–1560.

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nat. Neurosci.*, *6*(3), 309–315.

Spear, L. P. (2007). The developing brain and adolescent-typical behavior patterns: An evolutionary approach. In D. W. Romer & E. Walker (Eds.), *Adolescent psychopathology and the developing brain: Integrating brain and prevention science* (pp. 9–30). New York: Oxford University Press.

Sternberg, R. J., & Downing, C. J. (1982). The development of higher-order reasoning in adolescence. *Child Dev.*, *53*, 209–221.

Sternberg, R. J., & Nigro, G. (1980). Developmental patterns in the solution of verbal analogies. *Child Dev.*, *51*, 27–38.

Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. *J. Exp. Child Psychol.*, *27*(2), 195–232.

Tamm, L., Menon, V., & Reiss, A. L. (2002). Maturation of brain function associated with response inhibition. *J. Am. Acad. Child Adolesc. Psychiatry*, *41*(10), 1231–1238.

Thomas, K. M., King, S. W., Franzen, P. L., Welsh, T. F., Berkowitz, A. L., Noll, D. C., Birmaher, V., & Casey, B. J. (1999). A developmental functional MRI study of spatial working memory. *NeuroImage*, *10*(3, Pt. 1), 327–338.

Vaidya, C. J., Bunge, S. A., Dudukovic, N. M., Zalecki, C. A., Elliott, G. R., & Gabrieli, J. D. (2005). Altered neural substrates of cognitive control in childhood ADHD: Evidence from functional magnetic resonance imaging. *Am. J. Psychiatry*, *162*(9), 1605–1613.

van den Wildenberg, W. P. M., & van der Molen, M. W. (2004). Developmental trends in simple and selective inhibition of compatible and incompatible responses. *J. Exp. Child Psychol.*, *87*(3), 201–220.

Velanova, K., Wheeler, M. E., & Luna, B. (2008). Maturational changes in anterior cingulate and frontoparietal recruitment support the development of error processing and inhibitory control. *Cereb. Cortex*, *18*(11), 2505–2522.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., Menezes Santoa, M. de, Thomas, C. R., & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychol. Sci.*, *10*(2), 119–125.

Wechsler, D., & Stone, C. P. (1945). *Wechsler Memory Scale*. New York: Psychological Corporation.

Welsh, K., Butters, N., Hughes, J., Mohs, R., & Heyman, A. (1991). Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. *Arch. Neurol.*, *48*, 278–281.

Wendelken, C. B., & Bunge, S. A. (under review). Relational processing: Distinct contributions of anterior prefrontal cortex and the hippocampus. *Nat. Neurosci.*

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Front. Hum. Neurosci.*, *1*(8).

Zelazo, P. D., Craik, F. I., & Booth, L. (2004). Executive function across the life span. *Acta Psychol. (Amst.)*, *115*(2–3), 167–183.

II

PLASTICITY

Chapter 6 HORNG AND SUR 91

7 WHITLOCK AND MOSER 109

8 LI AND GILBERT 129

9 PASCUAL-LEONE 141

10 BAVELIER, GREEN, AND DYE 153

11 STEVENS AND NEVILLE 165

Introduction

HELEN NEVILLE AND MRIGANKA SUR

PLASTICITY—THE ABILITY of the brain to change adaptively during learning and memory or in response to changes in the environment—is one of the most remarkable features of higher brain function. Five years ago, in the third edition of *The Cognitive Neurosciences*, researchers described many new mechanisms at multiple levels of the neuraxis that produce and regulate neuroplasticity. These included the genesis of new neurons and glia throughout life, and the role of adult stem cells in plasticity. The papers comprising the plasticity section in 2004 were all conducted in non-human animals. Over the last five years animal studies of neuroplasticity have continued apace. Additionally, a burgeoning literature on human neuroplasticity has emerged. The papers in the current section on plasticity describe both animal and human research and have several themes in common. Several point to the key role of attention in neuroplasticity and also to the malleable nature of attention itself. Several also describe the two sides of plasticity: the systems that are most changeable by environmental input are both more enhanceable and more vulnerable to deficit. Another theme that was repeated and that is important to reiterate is that different functions and related brain systems display different degrees and time periods of maximal plasticity, whereas others display equivalent plasticity throughout life. An exciting new development is the recognition of the key role that genes and molecules play in neuroplasticity. Allelic variation within several specific genes is a major determinant of the degree to which neuroplasticity is evident in both animals and humans. Furthermore, plasticity is manifested through molecular mechanisms that transduce electrical activity in the brain into changes in the weights of synapses or into patterns of synaptic contact and network connections. While these mechanisms have been studied extensively in animal models of developmental or adult

plasticity, they have clear implications for understanding neuroplasticity in the human brain.

Topographic projections, or maps, are fundamental for representing and analyzing sensory information in the brain. In chapter 6, Horng and Sur describe guidance and patterning molecules that underlie the formation of retinotopic maps in the visual pathway. Such maps form a scaffold of connections that is subsequently refined by activity-dependent plasticity. Target molecules themselves can be altered to induce "rewiring" of inputs from the retina to the auditory pathway in ferrets and mice. The auditory cortex, when driven by vision, develops key features of the visual cortex such as visual-orientation-selective cells and maps, demonstrating that the nature of input activity during development is crucial for creating networks that process the input. Finally, the projection from the two eyes to primary visual cortex forms another model system in which molecular mechanisms that refine cortical connections are being rapidly discovered.

In chapter 7, Whitlock and Moser describe mechanisms of plasticity in a brain region that is critical for the formation of episodic memory and that has been studied extensively in the adult animal brain: the CA1 region of the rat hippocampus. They describe the link between long-term synaptic potentiation and behavior, and show how synaptic plasticity supports the formation of place cells in the rat's hippocampus. More generally, place cells are part of neural networks which have different states that can aid in the storage or recall of representations or memories.

In chapter 8, Li and Gilbert describe plasticity in the primary visual cortex of adult primates as a correlate of perceptual learning. Even early sensory cortical areas retain the capacity for synaptic and network changes, so that repeated perceptual experiences and familiarity elicit enhanced sensitivity to specific stimulus features. Such plasticity often involves an interplay between top-down and bottom-up processing, such that neuronal responses are often dynamically influenced by the nature of the task being performed or the context in which a stimulus appears. The efficient encoding of behaviorally relevant stimuli within a dynamic cortical network is a manifestation of the plasticity seen in perceptual learning.

In chapter 9, Pasqual-Leone describes neuroplasticity in studies of sighted and blind individuals and those who have sustained brain damage. Employing transcranial magnetic stimulation of different frequencies, he can enhance and decrease neuroplasticity. Furthermore these studies can differentiate the neural changes that are necessary for improvement in behavior from those changes that are not, and from changes that instead are harmful. He also describes several genetic polymorphisms that constrain the neuroplasticity in these populations.

In chapter 10, Bavelier, Green, and Dye describe the remarkable enhancements in several cognitive functions in individuals who engage in fast-action video games. They propose that such activity results in Bayesian learning—that is, enhanced learning in the course of optimizing the rewards associated with success in video gaming. They discuss the different ways that harnessing the factors that are important in these effects could benefit educational programs and performance in the workplace.

In chapter 11, Stevens and Neville describe the different profiles of neuroplasticity and the two sides of neuroplasticity in the human visual, auditory, language, and attentional systems. These studies have been conducted on individuals with visual or auditory deprivation and of children of different ages. They also describe their recent studies of interventions that target the most plastic and vulnerable brain systems in children with or at risk for neurocognitive deficits.

6 Patterning and Plasticity of Maps in the Mammalian Visual Pathway

SAM HORNG AND MRIGANKA SUR

ABSTRACT Maps at successive stages of the visual system, and in particular visual cortex, organize salient stimulus features into complex cortical networks. Retinotopic maps and ocular dominance domains arise during development using a molecular program that specifies the rough topographic order of projections. Genetic mutations in mice have identified guidance and patterning cues that mediate this organization of maps and may lead to the creation of new maps. Spontaneous activity produced in the retina refines the precision of the maps before eye opening, and patterned activity after eye opening drives further refinement and maintenance. For ocular dominance, the cortex has a critical period for synaptic plasticity during which it is especially responsive to changes in input. During this time, changes in eye-specific drive lead to Hebbian and homeostatic changes in the cortical network. This potential for plasticity represents a functional reorganization in response to changing demands from the outside world and allows the organism to adapt to its environment.

A critical function of the brain is to provide an orderly and efficient neural representation of salient sensory stimuli from the outside world. In the mammalian visual pathway, representations of light reflectance in visual space are relayed from the retina as a topographic map to the thalamus and superior colliculus. Along this pathway, projections from the two eyes are kept in parallel. Retinotopic and eye-specific information from the thalamus are transferred to the primary visual cortex, where additional stimulus features are extracted. The mechanisms by which visual stimulus feature maps are established and modified in response to experience are an active area of research, as these mechanisms are central to specifying the organizational details of the visual pathway and the functional characteristics of vision.

In this chapter, we will review the processes of retinotopic mapping and cortical plasticity in the mammalian brain. Molecular mechanisms of these phenomena have been studied most extensively in the mouse, a model for which genetic manipulations are available. What we currently know of these mechanisms illustrates how circuits are shaped by genetic programs, electrical activity, and experience-dependent modulation of stimulus input.

During development, the formation of a retinotopic map requires that axons responsive to neighboring positions in visual space maintain their relative positions as they innervate their target. This process involves graded patterns of guidance receptors expressed across the population of axons and matched to a complementary gradient of ligands on the target cells. The genetic patterning of guidance cues confers a rough order and spatial efficiency to the retinotopic map. However, further retinotopic precision and ocular dominance segregation depend upon patterns of spontaneous activity in the retina and experience-driven input. Changes in the level or pattern of activity can alter the structure and function of the retinotopic map in early development and of ocular dominance regions in later development and adulthood. Ocular dominance plasticity occurs in response to changes in competitive input between the eyes, and a variety of molecular pathways, many of which reflect the maturational state of the circuit, have been implicated in this process. Thus the developmental context shapes the extent to which changes induced by competitive input between the eyes occur. Here, genetic programs of development interact with activity- and experience-dependent input to mediate map refinement and plasticity.

The formation of the visual pathway during early development

Regionalization of Visual Pathway Centers

Functional pathways of the brain arise out of genetic programs of early development, which establish structural regions and wire them together (Rakic, 1988; O'Leary, 1989; Job & Tan, 2003; Sur & Rubenstein, 2005). During embryogenesis, sources of diffusible molecules, called signaling centers, induce regional and graded patterns of gene expression in the anterior neural tube. These patterns translate into structurally parcellated and functionally differentiated brain regions, including those devoted to processing incoming visual stimuli (Figdor & Stern, 1993; Rubenstein, Martinez, Shimamura, & Puelles, 1994; Rubenstein, Shimamura, Martinez, & Puelles, 1998;

SAM HORNG and MRIGANKA SUR Department of Brain and Cognitive Sciences, The Picower Institute for Learning and Memory, MIT, Cambridge, Massachusetts

Ragsdale & Grove, 2001; Nakagawa & O'Leary, 2002; Grove & Fukuchi-Shimogori, 2003; Shimogori, Banuchi, Ng, Strauss, & Grove, 2004).

In the mouse, centers of the visual pathway are established in this way: neuromeres P2 and P3 of the diencephalon differentiate into the dorsal and ventral thalamus, respectively, and around E13.5, lateral nuclei cluster to form the dorsal and ventral subdivisions of the lateral geniculate nucleus (LGN) (Jones, 1985; Tuttle, Braisted, Richards, & O'Leary, 1998). From E11 to E19, area 17 of posterior cortex differentiates in response to cortical gradients of FGF8, Wnts, BMP, and Shh to form the primary visual cortex (V1) (Dehay & Kennedy, 2007). How continuous gradients of gene expression throughout the neural tube are translated into boundary-delimited regions of functionally specific identities is not yet known; moreover, region-specific gene expression has not yet been reported (Nakagawa & O'Leary, 2001; Jones & Rubenstein, 2004).

Targeting and Retinotopic Wiring The visual pathway is wired (figure 6.1*A*) when roughly one-third of the ganglion cell axons from the retina project to the dorsal and ventral subdivisions (LGNd, LGNv) of the LGN while the remaining two-thirds target the superior colliculus (SC) in the brain stem (Jones, 1985; Tuttle et al., 1998). Axonal pathfinding to fugal (i.e., thalamic) and collicular targets begins around E15–16 and peaks at E19 (Colello & Guillery, 1990; Figdor & Stern, 1993; Tuttle et al., 1998; Inoue et al., 2000; Gurung & Fritzsch, 2004; Guido, 2008). In the mouse, axons from the ventrotemporal retina project ipsilaterally while the rest of the axons project contralaterally, with contralateral innervation to the thalamus occurring earlier (E15–16) than ipsilateral targeting (P0–2) (Dräger & Olsen, 1980; Godement, Salaun, & Imbert, 1984). Connections between the LGN (in this review, LGN is used to denote LGNd) and V1, both in the feedforward geniculocortical direction and in the feedback corticogeniculate pathway, emerge around E14 (Zhou et al., 2003).

Elucidating the mechanisms of retinotopic targeting and mapping has become a comprehensive field of study (figure 6.1*B*). Molecular mechanisms of retinal ganglion cell (RGC) guidance have been extensively studied in Xenopus, zebrafish, chick, and mouse models. Much of this work has focused on guidance to the optic disk, dessucation at the optic chiasm, and topographic map formation at the optic tectum, or SC (Inatani, 2005; Mann, Harris, & Holt, 2004). In the retina, laminin and netrin repulse DCC receptor-expressing RGC axons out of the optic head and into the optic nerve (Hopker, Shewan, Tessier-Lavigne, Poo, & Holt, 1999). Along the optic nerve, a repulsive semaphorin 5a sheath maintains the integrity of an interior axon pathway (Shewan, Dwivedy, Anderson, & Holt, 2002; Oster, Bodecker, He, & Sretavan, 2003). Slit1- and slit2-expressing cells guide repulsed robo-expressing axons to the proper decussation site for optic chiasm formation (Erskine et al., 2000; Ringstedt et al., 2000; Plump et al., 2002), and ephrin-B2 expression at the optic chiasm steers EphB1-receptor-expressing ventrotemporal axons ipsilaterally (Williams et al., 2003; Lee, Petros, & Mason, 2008). Matrix metalloproteinases (MMP) have been implicated in optic chiasm crossing and tectal targeting (Hehr, Hocking, & McFarlane, 2005). Additionally, traditional morphogens influence retinal ganglion cell pathfinding (Charron & Tessier-Lavigne, 2005): FGF-2 repels RGC growth cones along the optic tract (Webber, Hyakutake, & McFarlane, 2003), BMP7 promotes axonal outgrowth at the optic disk (Carri, Bengtsson, Charette, & Ebendal, 1998; Bovolenta, 2005), and Shh exhibits concentration-dependent attractive or repulsive effects in the retina and optic chiasm, respectively (Trousse, Marti, Gruss, Torres, & Bovolenta, 2001; Kolpak, Zhang, & Bao, 2005). Less is known about the specific cues mediating ganglion cell ingrowth to the LGN and geniculocortical targeting to area 17, or V1.

However, molecules that contribute to the topographic ordering of projections have been investigated in the LGN and V1, as well as the SC. The spatial position of visual stimuli is inverted through the lens and encoded on a sheet of retinal ganglion cells. This topographic map gets projected into the LGN and V1, as well as the SC. Because axon guidance cues must not only flag targets but also confer information about the relative topography of neighboring axons, positional cues are needed to maintain the retinotopic order of the projecting pathway. To avoid employing an infinitely large number of distinct positional cues, a gradient of one molecule along the sheet of axons may be matched to a complementary gradient of its binding partner in the target (Sperry, 1963). This "chemoaffinity" model has been confirmed with the discovery of a number of different receptor-ligand gradients expressed in projecting axons and target cells along the visual pathway.

The most comprehensively studied of these graded mapping molecules are the ephrin ligands and Eph family of tyrosine kinase receptors (figure 6.1*B*). The contribution of ephrinA-EphA receptor interactions to topographic mapping was first described in the optic tectum, where low-to-high ephrin-A2/A5 expression along the anterior-posterior axis was found to interact with a complimentary high-to-low EphA3 receptor gradient in terminals of the temporal-nasal axis of retina (Nakamoto et al., 1996; Feldheim et al., 1998, 2000; Hansen, Dallal, & Flanagan, 2004; Bolz et al., 2004). Interactions between ephrin-A and EphA receptors were initially thought to be repulsive, though subsequent studies revealed a concentration-dependent transition from attraction to repulsion: with low ephrin-A concentrations causing axonal attraction and high levels causing repulsion (Hansen et al., 2004). The ability of one ligand-receptor system to both attract and repel allows for

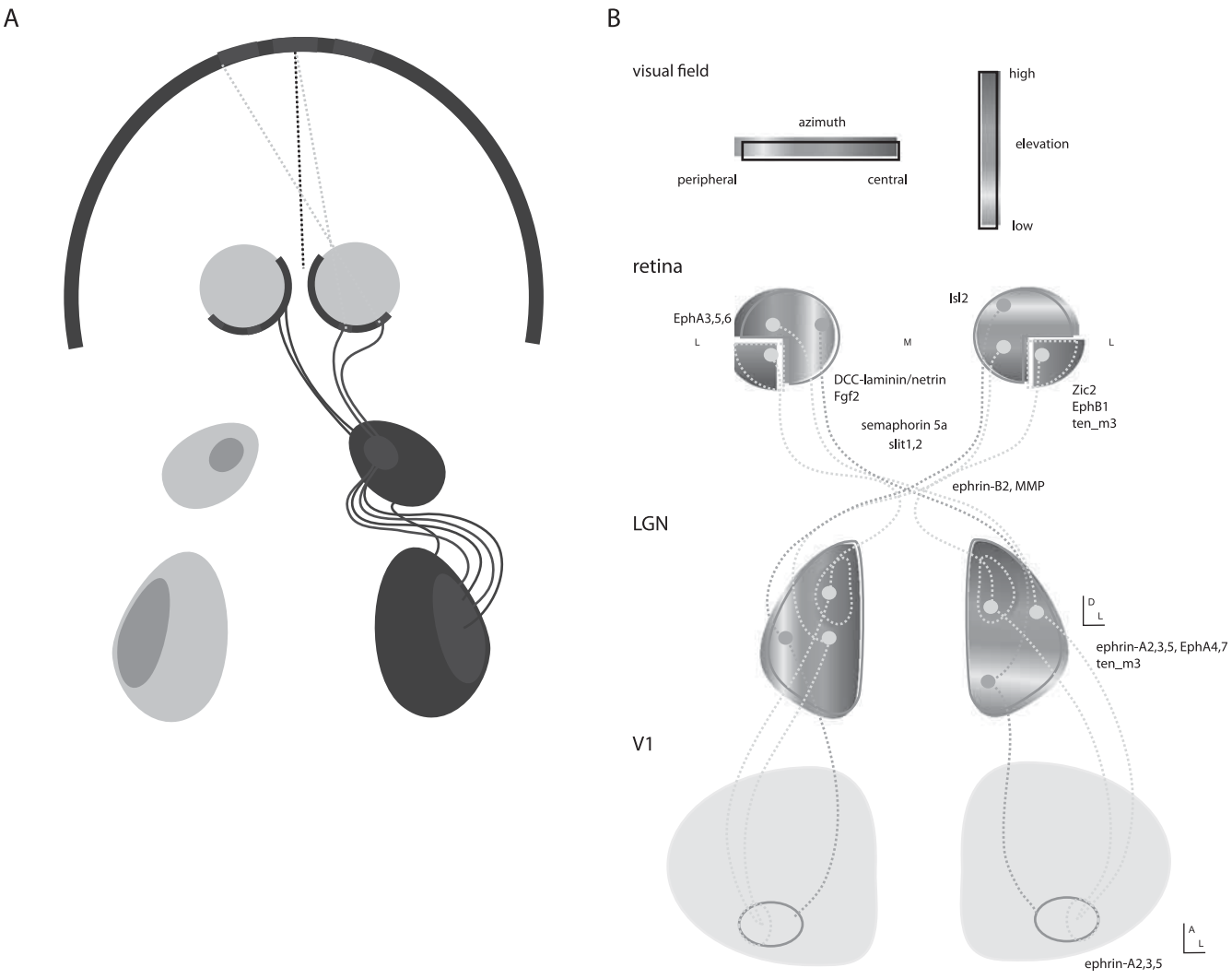


FIGURE 6.1 (*A*) Representation of the rodent visual pathway. Retinal ganglion cells project to the LGN, which in turn projects to the primary visual cortex (V1). A central region of the visual field is represented by both eyes along the pathway (ipsilateral, red; contralateral, blue). Contralateral and ipsilateral retinal ganglion cell terminals representing this binocular region are segregated in the LGN (red, ipsilateral zone; blue, contralateral zone). Geniculocortical fibers representing this region converge onto a binocular zone located in the lateral half of V1 (red, binocular zone; blue, monocular zone). (*B*) Schematic representation illustrating retinotopic map organization at each stage of the visual pathway and known guidance cues contributing to patterning. The visual field can be divided into two Cartesian axes, azimuth and elevation. For clarity, the azimuthal map on the left is diagrammed onto the visual pathway of the right hemisphere. The elevation map on the right is diagrammed onto the pathway of the left hemisphere. In reality, both axes of visual space are represented concurrently in both hemispheres. The ganglion cell sheet of the retina is divided into a contralaterally projecting region and an ipsilaterally projecting region. The ipsilateral retina originates from the ventrotemporal quadrant and is characterized in late embryogenesis by Zic2 and EphB1 expression. Conversely, the contralateral retina is characterized by Isl2 expression. Retinal ganglion cells express DCC and are repulsed out of the optic head by laminin and netrin. Factors, such as semaphoring-5a, keep retinal axons on course in the optic tract, where ipsilateral axons are repulsed by ephrin-B2 while contralateral axons decussate. High temporal to low nasal gradients of EphA receptor and ten_m3 expression in retinal axons likely influence terminal zones onto gradients of ephrin-A in the LGN. Ipsilateral axons terminate in a dorsomedial core of the LGN, segregated from surrounding contralateral axons. Activity-dependent refinement is necessary for proper eye-specific segregation. While ephrin-A gradients shape retinotopic termination zones, ten_m3 specifically influences ipsilateral targeting. Geniculocortical axons innervate V1. Ipsilateral inputs and corresponding contralateral fibers converge in the lateral binocular zone, while contralateral inputs representing regions not detected by the ipsilateral eye terminate in the medial monocular zone. Loss of ephrin-As leads to the disorganization of cortical maps only on the azimuthal axis, suggesting that other, unidentified factors contribute to the mapping of elevation. (See color plate 2.)

the target to be filled more parsimoniously than with separate attractant and repulsant molecules.

High lateral-to-medial gradients of ephrin-A2/A5 are also present in the mouse and ferret LGN and direct topography of high levels of EphA5/A6 expression from the contralateral nasal projections and low levels in the ipsilateral temporal projections (Huberman, Murray, Warland, Feldheim, & Chapman, 2005). Loss and ectopic gain of ephrin-A2, 3 and 5 lead to disruptions of the topographic map in both the LGN and V1: loss produces a medial shift in V1, in addition to internal disorganization, while lateral overexpression leads to a compression of V1, suggesting that EphA-expressing geniculocortical axons respond to a high medial to low lateral gradient of ephrinA (Cang et al., 2005). High dorsal EphB receptor expression in retina responds to low ventral ephrin-B expression in the tectum, and EphB-ephrinB gradients are speculated to similarly organize distinct axes in the LGN and V1 (Hindges, McLaughlin, Genoud, Henkemeyer, & O'Leary, 2002; McLaughlin, Hindges, Yates, & O'Leary 2003). The role of potential cis and trans mediated interactions among ephrin and Eph receptors from countergradients expressed on axons of the same area have yet to be explored (Luo & Flanagan, 2007). Finally, additional graded positional cues have been identified in the retinotectal map. Repulsive guidance molecule (RGM), a novel membrane-associated glycoprotein expressed in the posterior tectum, repulses temporal axons in vitro (Monnier et al., 2002), while engrailed-2 (En-2), a homeodomain transcription factor, is secreted by the posterior tectum, is endocytosed into axons, and attracts nasal axons while repelling temporal axons (Brunet et al., 2005). A high-to-low gradient of Wnt3 in the medial-lateral axis of the optic tectum mediates patterning via ventral-dorsal differences in Ryk receptor expression (Schmitt et al., 2006), and a Wnt signaling inhibitor, SFRP1, interacts with RGC receptor Fz2, to steer axons along the optic tract en route to the tectum (Rodriguez et al., 2005).

Experiments in which half of retinal ganglion cells are ablated or a disordered set of cells gain EphA expression reveal that retinotectal axons persistently fill their target (Brown et al., 2000; Feldheim et al., 2000). Thus it is the relative level of positional information rather than absolute signaling that determines the topography of retinal axons. Some limiting factor, whether from the axon-axon interaction or target-derived cues, may ensure that target filling occurs (Luo & Flanagan, 2007). Loss of L1CAM leads to incomplete filling of the tectum, and this molecule may have such a role (Demyanenko & Maness, 2003).

Eye-Specific Domains A second fundamental organizational feature of the visual pathway is its segregation into eye-specific domains. Maintaining parallel channels for eye-specific input allows for stereoscopic vision, or depth perception. In mice, ipsilateral projections form a dorsal core in the LGN (LGNd) and are flanked laterally by contralateral terminals representing matched areas of visual space. These axons intermix when projecting to layer IV cells of the binocular zone, a V1 subregion bounded medially by a monocular zone of contralateral input (figure 6.1*A*,*B*). In mammals with more complex visual systems, such as the ferret, cat, primate, and human, eye-specific domains form a map of ocular dominance stripes in V1 (figure 6.3). Whether eye-specific domains are influenced by positional cues in addition to activity dependent processes of terminal segregation has only recently begun to be explored. Developmental time course studies in the mouse show that early (P0–P5) ipsilateral axons are diffusely targeted to the dorsal-medial portion of the LGN and progressively become more strictly confined to a central core by P28 (Jaubert-Miazza et al., 2005). Although activity-dependent processes to be discussed later contribute to the refinement of ocular domains in the LGN (Shatz, 1983; Shatz & Stryker, 1988; Pfeiffenberger et al., 2005), the initial ingrowth of ipsilateral axons shows a bias toward the binocular region in the central part of the dorsal half of the LGN, and eye-specific guidance cues likely instruct this initial positioning (Godement et al., 1984). The presence of functional markers, Isl2 and Zic2, during late embryogenesis (E13–E17), for contralaterally and ipsilaterally projecting retinal ganglion cells, respectively, suggests that the two populations have distinct differentiation programs and potentially respond to unique cues in their target (Herrera et al., 2003; Pak, Hindges, Lim, Pfaff, & O'Leary, 2004). Loss of ten_m3, a homophilic binding protein expressed strongly on ipsilaterally projecting axons, leads to the selective ventral expansion of ipsilateral axons and no disruption in contralateral axons in the LGN (Leamey, Glendining, et al., 2007; Leamey, Merlin, et al., 2007). Therefore, ten_m3 and potentially other unknown cues may contribute to the formation of eye-specific domains. Mechanisms of how corresponding ipsilateral and contralateral axons are coordinated and aligned to form binocular maps are poorly understood.

Other Feature Maps and the Formation of New Maps In mice and other mammals, additional stimulus features are encoded in the visual pathway at the cortical level. Cells in V1 are selective for orientation, spatial frequency, and the direction of visual stimuli. In carnivores and primates, these cells are organized into selectivity maps of their own. For example, multiple stripes converging around a pinwheel center on the cortical surface represent graded regions of different orientation selectivity. Within these orientation-selective regions, directionally selective subregions are present. Using a layout that maximizes map continuity and cortical coverage (Swindale, Shoham, Grinvald, Bonhoeffer, & Hübener, 2000), multiple feature

maps are superimposed and organized in systematic fashion, with regions of high gradients from different maps spatially segregated from one another (Yu, Farley, Jin, & Sur, 2005). That is, while individual, adjacent neurons respond best to different values of the same feature, the way in which these features are mapped varies systematically. The critical parameter is the rate of change of each feature across the same set of neurons: at locations where one feature changes rapidly, other features change little.

Mechanisms of map formation for these additional stimulus features are not well understood, although the role of intrinsic genetic programs of patterning and activity-dependent input may differ depending on the specific feature map (White & Fitzpatrick, 2007). Whereas the retinotopic and eye-specific maps are patterned roughly before birth and eye opening, the orientation map is detectable only by the time of eye opening in the ferret (Chapman, Stryker, & Bonhoeffer, 1996; White, Coppola, & Fitzpatrick, 2001; Coppola & White, 2004), and the direction-selective map appears 1–2 weeks later (Li, Fitzpatrick, & White, 2006). Therefore, the formation of these maps likely depends critically on developmental processes coincident with patterned input into the cortex.

The formation of orientation maps coincides with a period during which axonal connections in V1, especially long-range horizontal inhibitory projections in layer 2/3, proliferate (Bosking et al., 2002). Orientation tuning has been hypothesized to arise from feedforward patterns of thalamocortical connectivity (Ferster & Miller, 2000) and to be shaped by intracortical connections (Somers, Nelson, & Sur, 1995) and balanced inhibition (Marino et al., 2005). The maturation of this supragranular inhibitory network may contribute to the appearance of orientation tuning and organization of tuned cells into selective domains. Mice deficient in Arc, an activity-dependent cytoskeletal-associated protein implicated in the synapse-specific modulation of AMPA receptor number, show weaknesses in orientation tuning in V1 (Wang et al., 2006). Dark-reared animals exhibit a delay in the formation of the orientation map, while binocularly lid-sutured animals have a near complete degradation of the map (White et al., 2001), suggesting that low levels of non-patterned activity have a greater disruptive effect than the absence of input. Therefore, unknown intrinsic properties of the cortex instruct the formation of the orientation map in the weeks after eye opening and induce the map even in the absence of vision. However, the orientation map is susceptible to disruption in response to disorganized activity.

In contrast to orientation maps, a 2-week period following eye opening is both necessary and sufficient for the formation of direction-selective maps (Li et al., 2006). Thus the direction-selective map is induced by changes in either the cortex or LGN that are driven by activity. Sharpening of retinotopic tuning and decreases in the response latency of LGN cells may play a role (Tavazoie & Reid, 2000). Different feature maps in V1 appear to be guided by independent mechanisms. Loss of the direction-selective map leaves the orientation map intact, and monocular enucleation to eliminate the ocular dominance map does not interfere with the formation of the remaining V1 feature maps (Farley, Yu, Jin, & Sur, 2007). However, the relative positioning of different maps in V1 is responsive to alterations in a given map, as monocular enucleation leads to the coordinated reorganization of the remaining map dimensions (Farley et al., 2007). Therefore, while the formation of stimulus-specific maps or networks likely relies on unique developmental mechanisms, whether they be genetically determined or instructed by activity, the detailed organization of each map and its structural and spatial coordination with other maps is a key feature of activity-dependent cortical organization.

Because of the independent origin of individual maps (and response features), the appearance of new maps in evolution may have depended on unique events and developmental processes for a given map. However, there may be general properties in neural circuits that allow for the introduction of a novel map. Novel maps may arise potentially through the duplication and subsequent functional divergence of an existing map, or the addition of a novel input into an existing region and subsequent reorganization of cortical circuitry into a new map. An example of the former is the induction of duplicate barrel cortices by ectopic posterior cortical FGF8 expression (Fukuchi-Shimogori & Grove, 2001). An example of novel input leading to the introduction of a new map includes the implantation of a third eye leading to triple ocular dominance stripes in the tectum of the frog (Constantine-Paton & Law, 1978), rewired retinal input to the MGN driving retinotopic maps to form in primary auditory cortex (A1) (Sur, Garraghty, & Roe, 1988), and ten_m3 mutation in mouse leading to a medial expansion of ipsilateral input to V1 and the de novo formation of cortical ocular dominance stripes (C. Leamey, personal communication).

Rewiring vision into the auditory pathway

After neonatal surgical ablation of the inferior colliculus (IC), retinal ganglion cells are rerouted to target the auditory thalamus and subsequently induce the auditory pathway to process visual information (figure 6.2). This experimental paradigm allows us to investigate the role of novel input in producing retinotopic and feature maps, and to screen for unknown guidance cues involved in wiring together sensory pathways. The normal auditory pathway comprises cochlear afferents projecting to the inferior colliculus (IC), which sends fibers along the brachium of the IC (BIC) to the medial geniculate nucleus (MGN) in the thalamus, which then innervates the primary auditory cortex (A1; figure 6.2*A*). Using hamsters, Schneider discovered that retinal afferents form

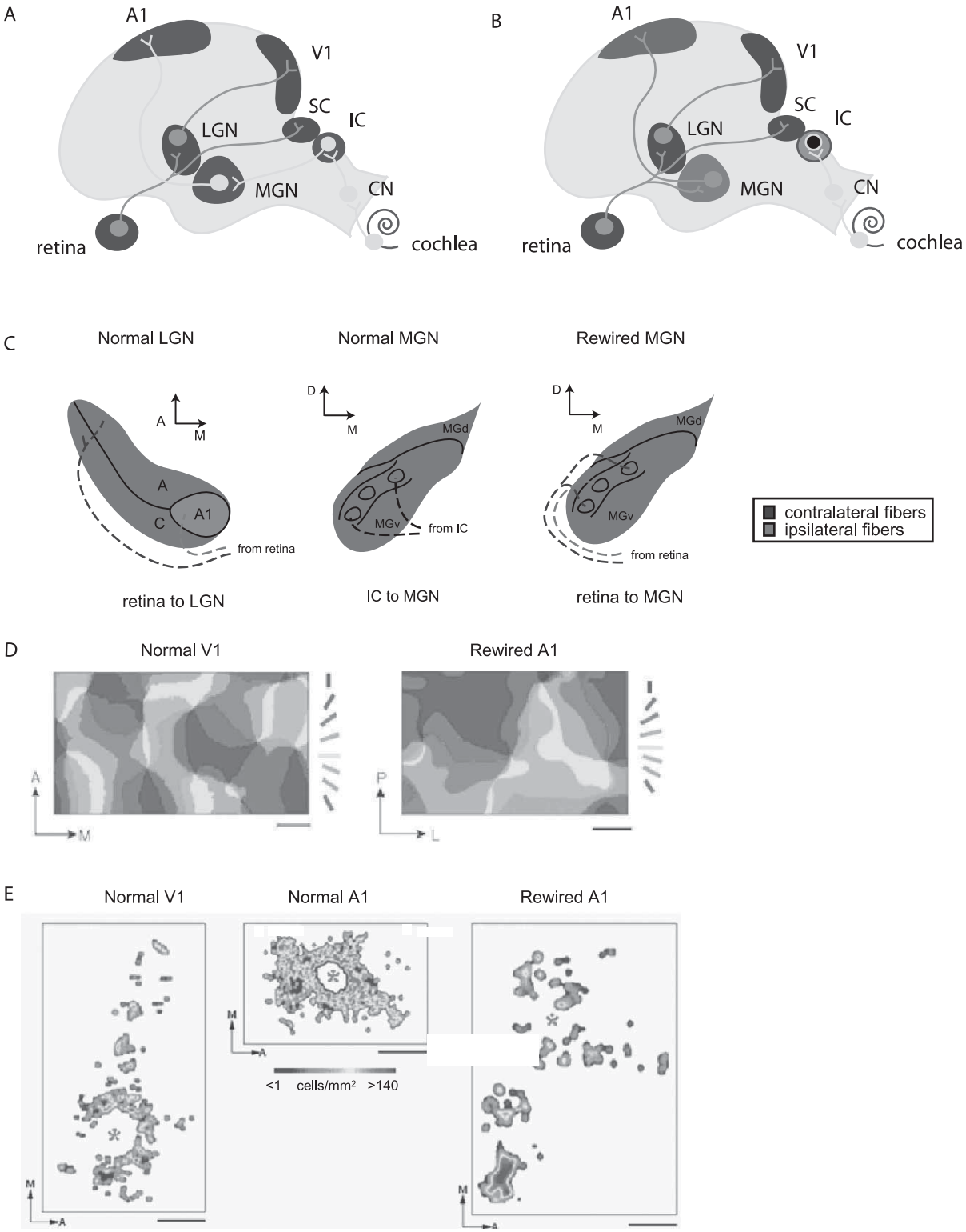

A
A1
V1
SC
IC
LGN
MGN
CN
retina
cochlea
B
A1
V1
SC
IC
LGN
MGN
CN
retina
cochlea
C
Normal LGN
Normal MGN
Rewired MGN
A
M
A
A1
C
from retina
D
M
MGd
MGv
from IC
D
M
MGd
MGv
from retina
contralateral fibers
ipsilateral fibers
retina to LGN
IC to MGN
retina to MGN
D
Normal V1
Rewired A1
A
M
P
L
E
Normal V1
Normal A1
Rewired A1
M
A
<1 cells/mm² >140
M
A
M
A

FIGURE 6.2 Primary visual and auditory pathways in normal and rewired mice: anatomical and physiological consequences of rewiring. (*A*) The visual pathway in ferrets and mice begins with retinal projections to the lateral geniculate nucleus (LGN) and superior colliculus (SC). The LGN projects in turn to the primary visual cortex (V1). The auditory pathway traces from the cochlea to the cochlear nucleus (CN) and then to the inferior colliculus (IC). From IC, connections are made with the medial geniculate nucleus (MGN), which projects to the primary auditory cortex (A1). (*B*) Ablation of the IC in neonatal animals induces retinal afferents to innervate the MGN and drive the auditory cortex to process visual information. (*C*) Retinogeniculate axons of normal ferrets project to eye-specific regions of the LGN (horizontal plane), while IC afferents project to the ventral subdivision (MGv) of the MGN (coronal plane) and innervate lamellae parallel to the lateral-medial axis. Rewired auditory fibers innervate the MGv along adjacent, nonoverlapping eye-specific terminals within MGv lamellae. (Adapted from Sur & Leamey, 2001.) (*D*) Orientation maps are present in normal V1 and rewired A1 of ferrets using optical imaging of intrinsic signals. The animal is stimulated with gratings of different orientations, while hemodynamic changes in red wavelength light reflectance caused by increases in oxygen consumption are detected from the cortex with a digital camera. The orientation preference map is calculated by computing a vector average of the response signal at each pixel. Color bar: color coding representing different orientations. Scale bar: 0.5 mm. (*E*) Retrograde tracers reveal the pattern of horizontal connections in superficial layers of normal V1, normal A1, and rewired A1 of ferrets. Distribution of horizontal connections in rewired A1 more closely resembles that of normal V1 than normal A1 and potentially contributes to the refinement of orientation mapping in rewired A1. Scale bars: 500 μm. (Adapted from Sharma, Angelucci, & Sur, 2000.) (See color plate 3.)

novel connections to the ventral MGN (MGv) when the IC is ablated after birth (figure 6.2*B*; Schneider, 1973; Kalil & Schneider, 1975; Frost, 1982; Frost & Metin, 1985). This "rewiring" paradigm has subsequently been demonstrated and studied in the ferret and mouse models (Sur et al., 1988; Roe, Pallas, Hahm, & Sur, 1990; Roe, Pallas, Kwon, & Sur, 1992; Lyckman et al., 2001; Newton, Ellsworth, Miyakawa, Tonegawa, & Sur, 2004; Ellsworth, Lyckman, Feldheim, Flanagan, & Sur, 2005).

On receiving retinal ganglion cell input, the MGN adopts some of the anatomic and physiologic features of the normal LGN (figure 6.2*C*). Rewired MGN neurons of the ferret exhibit center-surround visual receptive fields (Roe, Garraghty, Esguerra, & Sur, 1993), topographic ordering (Roe, Hahm, & Sur, 1991), and eye-specific segregation (Angelucci, Clasca, Bricolo, Cramer, & Sur, 1997). The potential to form ordered retinotopic and ocular dominance regions in MGN indicates that common patterning cues exist between the LGN and MGN. Experiments in ephrin A2/A5 double knockout mice reveal that surgically induced rewiring is enhanced (Lyckman et al., 2001), with ipsilateral projections especially increased, as they originate from the temporal retina and express the highest levels of EphA receptor (Ellsworth et al., 2005). Loss of innervation to the MGN somehow makes this nucleus permissive to retinal axon ingrowth, and a gene-screening process between the normal and rewired MGN may facilitate the discovery of tropic or repulsive agents regulating retinal axon affinity for different sensory nuclei of the thalamus.

Nonetheless, certain morphological aspects of rewired MGN are resistant to change (figure 6.2*C*). In ferrets, retinal axon terminations are elongated along the typical isofrequency axis, or lamellae, of the MGN as opposed to more focal, isotropic distributions in the LGN (Pallas, Hahm, & Sur, 1994). In addition, eye-specific clusters are smaller and cruder than the eye-specific layers of LGN (Angelucci et al., 1997).

In the cortex of rewired ferrets, cells in A1 respond to visual field stimulation and form a functional retinotopic map of visual space (Roe et al., 1990). However, the thalamocortical axons transmitting this information retain their pattern of elongated projections along the anteroposterior axis of A1, which typically correspond to isofrequency bands (Pallas, Roe, & Sur, 1990). In order to create the functional map of focal retinotopic representations, either a refinement of these elongated inputs by a reorganized intracortical inhibitory network or a difference in drive along the projection itself is required (Sur, Pallas, & Roe, 1990). Consistent with the first possibility, calbindin-immunoreactive GABAergic neurons of rewired A1 have more elongated axonal arbors (Gao, Wormington, Newman, & Pallas, 2000). Thus, despite persistent structural features of A1 and thalamocortical input, functional retinotopy can be driven by novel patterns of activity.

In the ferret, rewired A1 acquires novel maps of orientation selectivity with pinwheels and orientation domains (figure 6.2*D*), similar in general to maps in normal V1 (Sharma et al., 2000; Rao, Toth, & Sur, 1997). In rewired A1, orientation maps are less organized, although intrinsic horizontal connections of superficial layer pyramidal neurons are clustered and bridge distantly located domains of the same orientation preference, as in V1 (figure 6.2*E*; Sharma et al., 2000). This pattern of intracortical connectivity is in contrast to horizontal connections in normal A1, where horizontal connections are limited to isofrequency domains of the tonotopic map and stretch along these bands. Such reorganization of horizontal connections driven by visual activity is likely related to changes in the inhibitory circuits of rewired A1, and it suggests that coordinated activity-dependent changes in inhibitory and excitatory networks of at least the superficial cortical layers are a prominent feature of cortical map organization and plasticity.

Finally, the rewired auditory pathway is sufficient to instruct visually mediated behavior. After training to

distinguish a left visual hemifield stimulus from an auditory stimulus, ferrets with a unilaterally rewired left hemisphere are able to accurately perceive a right visual hemifield stimulus as visual even after left LGN ablation (von Melchner, Pallas, & Sur, 2000). After left LGN ablation, the ferrets also possess diminished yet intact spatial acuity in the right hemifield. Subsequent ablation of the rewired A1 abolishes the animals' ability to distinguish a right hemifield stimulus presented as visual. Thus rewired A1 is sufficient and necessary in the absence of ipsilateral visual pathway input to detect a visual percept in trained ferrets. In mice, direct subcortical projections from the MGN to the amygdala are involved in rapid fear conditioning to an auditory cue (Rogan & LeDoux, 1995; Doran & LeDoux, 1999; Newton et al., 2004). Because of an indirect pathway from the LGN through V1 and the perirhinal cortex to the amygdala, a fear conditioning to a visual cue requires many more training sessions (Heldt, Sudin, Willott, & Falls, 2000). In rewired mice, the acquisition time of a fear conditioning to a visual cue is accelerated and resembles that of a normal mouse in response to an auditory cue (Newton et al., 2004).

Activity-dependent refinement of visual maps

Although topography of the retinotopic map and eye-specific domains are roughly established by programmed guidance and patterning cues, activity plays a critical role in the refinement and maturation of these maps. Single cells in the mouse LGN receive weak input from one to two dozen retinal ganglion cells, which occupy 30% of the cell surface, in the first postnatal week, and then begin to prune these connections down to one to three strong monocular inputs that occupy 1–5% of the cell surface (Chen & Regher, 2000; Jaubert-Miazza et al., 2005; Guido, 2008). Ipsilateral projections to the LGN are also diffuse and widespread during this first week, occupying nearly 60% of the nucleus area. By the time of eye opening (P12–P14), the ipsilateral zone occupies only 10% of the LGN (Jaubert-Miazza et al., 2005; Guido, 2008).

Both the retinotopic and eye-specific pruning of synapses is affected by altering spontaneous activity caused by cholinergic waves that sweep across the retina (Meister, Wong, Baylor, & Shatz, 1991; Wong, Meister, & Shatz, 1993). Blockade of retinal electrical activity with TTX (Harris, 1980) or loss of retinal waves by genetic loss of the β2 nAchR (Rossi et al., 2001; McLaughlin, Torborg, Feller, & O'Leary, 2003; Grubb, Rossi, Changeux, & Thompson, 2003; Chandrasekaran, Plas, Gonzalez, & Crair, 2005) causes terminals to remain desegregated and diffuse. Combined ephrinA and β2 nAchR mutants lead to additive defects in retinotopic organization in the LGN and V1 along the elevation axis in visual space, demonstrating that activity-dependent refinements form a strong contribution to the integrity of the retinotopic and eye-specific maps (Pfeiffenberger et al., 2005; Pfeiffenberger, Yamada, & Feldheim, 2006; Cang et al., 2008). The nob mutant mouse, which acquires an abnormal onset of high-frequency waves after eye opening, develops normal eye-specific segregation before eye opening, because early spontaneous waves are intact. After the onset of abnormal high-frequency waves, eye-specific inputs desegregate because of potentially synchronized firing between the eyes (Demas et al., 2006). Similarly, fish exposed to the synchronized stimuli of strobe illumination lose eye-specific segregation (Schmidt & Eisele, 1985). The most straightforward mechanism for these data involves the strengthening of correlated inputs and weakening and subsequent pruning of noncorrelated inputs (Hebb, 1949; Zhang & Poo, 2001). In retinogeniculate synapses, bidirectional changes in synaptic strength depend on the relative timing between optic tract stimulation and LGN depolarization (Butts, Kanold, & Shatz, 2007).

Specifically how decorrelated axon terminals are eliminated and persisting synapses strengthened is not well understood, though canonical immunologic signaling molecules may be involved in synaptic pruning. Loss of Class I MHC proteins, neuronal pentraxins, and the C1qb component of the complement cascade leads to persistently enlarged and desegregated ipsilateral zones in the LGN (Huh et al., 2000; Bjartmar et al., 2006; Stevens et al., 2007). These molecules are hypothesized to tag weak synapses for pruning during activity-dependent refinement. Notably, these manipulations do not affect the basic topographic organization of the retinotopic map and eye-specific domains. Conversely, ephrin-A mutants alone contain topographically disorganized, yet tightly refined, retinotectal terminals (Frisen et al., 1998; Feldheim et al., 2000).

In the mouse, retinotopic maps in V1 require patterned input for normal maturation. During the first 10 days after eye opening (P13–P23), normal activity brings eye-specific maps to adult levels of responsiveness and precision in receptive field organization (Smith & Tractenberg, 2007). The contralateral eye develops more precociously in map precision and magnitude, while the ipsilateral eye lags behind by roughly 5 days. When the contralateral eye is deprived, both contralateral and ipsilateral maps are delayed in retinotopic precision; when the contralateral eye is removed or silenced, precision of the ipsilateral maps is accelerated; when the contralateral eye is removed and ipsilateral eye deprived, the ipsilateral map precision is delayed. These data suggest that competing patterned inputs from both eyes is necessary for normal map refinement. Isolated patterned input accelerates map refinement, perhaps because of a lack of noise from the contralateral eye. The effects of binocular deprivation, or ipsilateral

removal plus contralateral deprivation, were not examined in this study.

In addition to Hebbian pruning and strengthening of feedforward inputs, changes due to activity that contribute to map refinement potentially involve additional developmental processes, including the remodeling of excitatory connections, the maturation of inhibitory circuits, and the timed expression of L-type Ca^{2+} channels. Excitatory synapses in the LGN initially contain NMDA receptors but increase their proportion of AMPA receptors as synaptic elimination proceeds (Chen & Regehr, 2000; X. Liu & Chen, 2008). Networks of GABAergic interneurons in the LGN also appear at P5 and mature by P14 (Ziburkus, Lo, & Guido, 2003; Jaubert-Miazza et al., 2005). L-type Ca^{2+} channels are expressed in excitatory LGN synapses before eye opening and are necessary for eye-specific segregation and CRE-mediated gene transcription (Cork, Namkung, Shin, & Mize, 2001; Pham, Rubenstein, Silva, Storm, & Stryker, 2001; Jaubert-Miazza et al., 2005).

In sum, mechanisms of map refinement in response to activity likely involve a number of different processes that contribute to the functional maturation of the circuit, including the selection and elimination of synapses, the modulation of synaptic strength, and the structural formation of inhibitory networks. Activity may also in turn influence the actions of guidance cues; activity blockade prevents ephrinA-mediated repulsion because of disruptions in cAMP signaling (Nicol et al., 2007).

Ocular dominance plasticity

Once a functional map is refined, ongoing patterns of activity contribute to the maintenance and alteration of this map in response to experience. The potential for plasticity is of particular interest for understanding how neural circuits adapt their structure and function to accommodate changing demands in the environment. Plasticity in the V1 ocular dominance map has become a paradigmatic model of activity-driven reorganization in network structure and function (figure 6.3*A*).

Structural and Functional Changes in Response to Lid Suture Within the binocular zone of V1 in mammals, neurons particularly in the superficial and deep

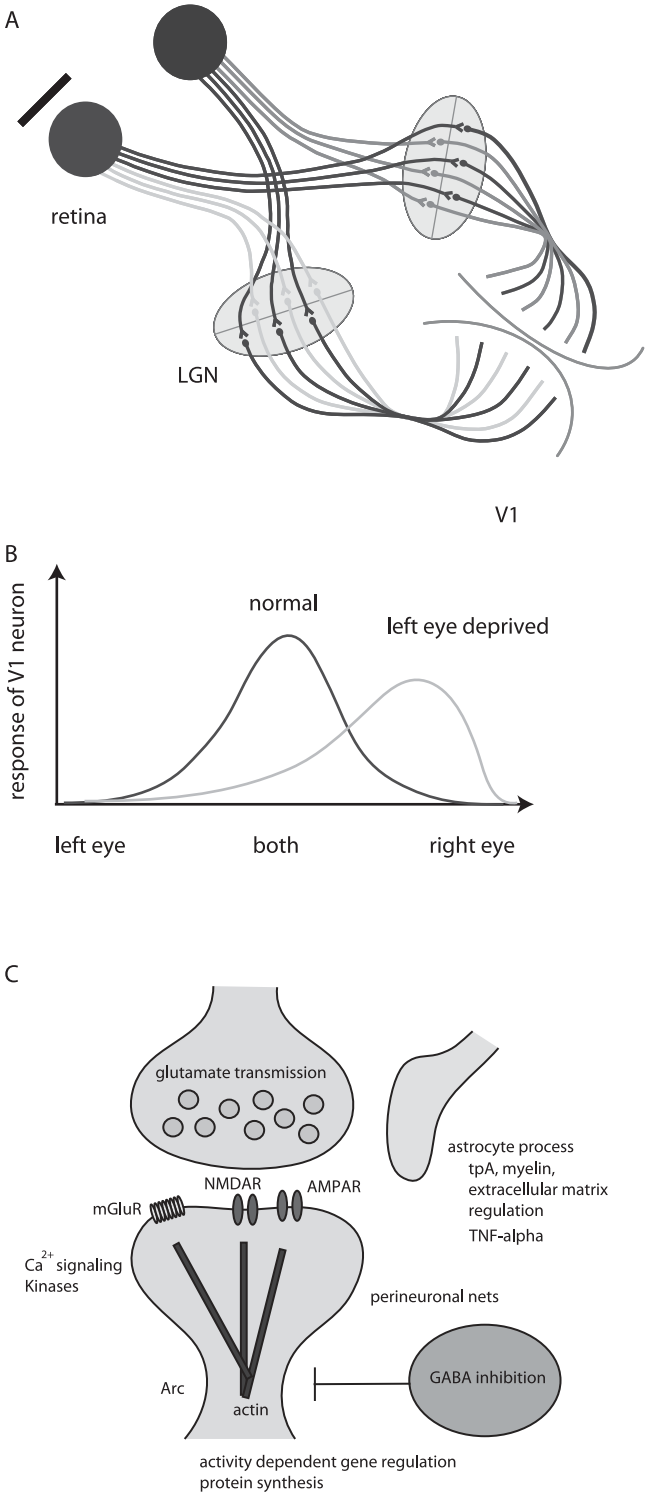


Figure 6.3 Ocular dominance anatomy and plasticity in V1. (*A*) Contralateral and ipsilateral fibers are segregated in the LGN but converge onto binocular cells in V1. (*B*) When one eye is deprived of input for a brief period during the critical period, or for a longer period during adulthood, binocular cells in V1 become more strongly driven by the nondeprived eye. Ocular dominance plasticity reflects both structural and functional changes of synapses. (*C*) The cellular and molecular mechanisms of ocular dominance plasticity are an active area of investigation. Processes known to play a critical role include signal transduction pathways downstream of the mGluR and NMDARs, and activity-dependent changes in AMPAR content at synapses, mRNA transcription, and protein translation. GABAergic inhibition is involved in inducing the critical period of ocular dominance plasticity, and extracellular matrix factors and perineuronal nets surrounding inhibitory interneurons have been implicated in constraining plasticity. (See color plate 4.)

layers of cortex are driven by both eyes, though neurons in layer 4 of carnivores and primates are primarily driven by one eye (Hubel & Wiesel, 1963; Stryker & Harris, 1986). When an imbalance of input occurs after lid suturing one eye for several days (monocular deprivation, or MD), a series of structural and functional changes leads to the weakening of the deprived eye input and the strengthening of nondeprived eye input (figure 6.3*B*). The mechanisms underlying these changes (figure 6.3*C*) shed light on core principles of plasticity in the developing brain in response to experience. Before functional shifts are apparent, spine motility increases (Majewska & Sur, 2003; Oray, Majewska, & Sur, 2004), followed by transient pruning of spines (Mataga, Mizuguchi, & Hensch, 2004; Oray et al.). Electrophysiological and optical imaging techniques reveal that deprived-eye connections are weakened first, while supragranular horizontal connections are remodeled (Trachtenberg, Trepel, & Stryker, 2000; Trachtenberg & Stryker, 2001; W. Lee et al., 2006). Strengthening of nondeprived eye connections follows (Frenkel & Bear, 2004), and finally, layer IV geniculocortical axons representing the nondeprived eye grow and expand their terminals at the expense of shrinking deprived-eye terminals (Antonioni & Stryker, 1996; Antonioni, Fagiolini, & Stryker, 1999). The chronology of these events has been best characterized for the developmental "critical period" in mouse, though differences in structural and physiological response may exist for MD during adulthood or under different paradigms of development, such as dark rearing (Jiang, Treviño, & Kirkwood, 2007). The developmental context under which MD is applied can make a qualitative and quantitative difference in the ocular dominance plasticity observed and likely involves different cellular and network mechanisms.

Critical Periods and the Developmental Context of Plasticity The ability to induce and reverse ocular dominance plasticity was initially thought to exist only during a "critical period" in development, a time approximately 10 days after eye opening during which short MD (a few days in the mouse) leads to a robust shift in ocular dominance toward the nondeprived eye (Hubel & Wiesel, 1970; Gordon & Stryker, 1996). This "critical period" is delayed by roughly three weeks in dark-reared animals (Cynader, Berman, & Hein, 1976; Fagiolini, Pizzorusso, Berardi, Domenici, & Maffei, 1994; G. Mower, 1991), suggesting that the cortex must reach a maturational state that is facilitated by a period of patterned vision. This maturational state has been shown to involve the development of an inhibitory network that depends on BDNF produced in response to neural drive (Hensch, 2005). Overexpression of BDNF leads to a precocious start of the critical period and premature development of inhibitory cells in the cortex (Huang et al., 1999; Hanover, Huang, Tonegawa, & Stryker, 1999), while administration of BDNF to dark-reared animals leads to the induction of a critical period (Gianfranceschi et al., 2003). Mice lacking polysialic acid also experience premature maturation of inhibitory networks and a precocious critical period (Di Cristo et al., 2007). GAD65 knockout mice, which lack axonal GABA synthesis and subsequent inhibitory transmission, do not experience a critical period unless induced with benzodiazepine drug infusion at any age (Hensch et al., 1998; Fagiolini & Hensch, 2000). Therefore, tonic GABA release is sufficient to mature an intracortical inhibitory network and induce critical period plasticity. The process by which the critical period closes and why critical period induction is a one-time event are not understood.

Although the critical period occurs once during development, longer periods of MD (7 to 10 days in the mouse) are able to trigger ocular dominance plasticity in adulthood (Sawtell et al., 2003; Hofer, Mrsic-Flogel, Bonhoeffer, & Hübener, 2006; Fischer, Aleem, Zhou, & Pham, 2007). This form of plasticity is thought to differ mechanistically from that experienced during the critical period, as nondeprived-eye connections are strengthened more rapidly and deprived-eye connections remain stable (Kaneko, Stellwagen, Malenka, & Stryker, 2008). Previous experiences with MD, either in the critical period or in adulthood, facilitate plasticity in response to short MD later in life (Hofer et al.; Frenkel & Bear, 2004), suggesting that a functionally suppressed anatomical trace has been laid. Ocular dominance plasticity may also be induced in adulthood after a 10-day period of visual deprivation, and this process mimics the time course of plasticity present during the critical period (He, Hodos, & Quinlan, 2006; Frenkel & Bear, 2004). Therefore, even the apparent closure of critical period plasticity may be reactivated by a brief loss of visual input.

Hebbian and Homeostatic Mechanisms of Plasticity Spike-timing-dependent activity has been demonstrated to lead to strengthening or weakening of geniculocortical and intracortical synapses in V1 (Frégnac & Shulz, 1999; Meliza & Dan, 2006). Long-term depression (LTD) of deprived-eye inputs occurs in vivo after MD and has been proposed to precipitate the eventual reduction in deprived-eye synapses (Heynen et al., 2003; Frenkel & Bear, 2004). Decreases in the threshold for LTD after dark rearing (as a result of decreases in NR2A/NR2B ratio of subunit composition in NMDA receptors) are posited to mediate the reactivation of plasticity (He et al., 2006). In hippocampal neurons, AMPA receptors are added to synapses during long-term potentiation (LTP) and removed during LTD (Manilow & Malenka, 2002), a mechanism that may act as the substrate for altering synaptic strength in visual cortex due to altered visual experience. Group 1 metabotropic glutamate receptors have been identified as inducers of LTD, and loss of mGlur5 blocks ocular dominance plasticity

(Dölen et al., 2007). Gene transcription and protein synthesis downstream of synaptic events is necessary for ocular dominance plasticity. Blocking cortical protein synthesis while preserving LTD effectively prevents ocular dominance shifts (Taha & Stryker, 2002). Loss of the cAMP responsive element, CREB, a protein that promotes CRE-mediated gene transcription, prevents ocular dominance shifts (A. Mower, Liao, Nestler, Neve, & Ramoa, 2002). Upstream Ca^{2+}-sensitive signaling kinases, including ERK, PKA, and CamKIIα, are also necessary for OD plasticity, and these likely activate a number of functional cascades that lead to gene transcription and structural modifications to the synapse (Di Cristo et al., 2001; Taha & Stryker, 2002; Berardi, Pizzorusso, Ratto, & Maffei, 2003; Suzuki, al-Noori, Butt, & Pham, 2004; Gomez, Alam, Smith, Horne, & Dell'Acqua, 2002; Chierzi, Ratto, Verma, & Fawcett 2005; Taha & Stryker, 2005). The extent to which LTD and LTP are necessary for ocular dominance shifts is uncertain, however. In mice lacking the protein phosphatase, calcineurin, LTD is blocked, but ocular dominance plasticity remains intact (Yang et al., 2005).

Although Hebbian mechanisms are likely to contribute to ocular dominance plasticity in which poorly driven synapses from the deprived eye are pruned and synapses from the nondeprived eye are strengthened (Katz & Shatz, 1996), additional cellular and network mechanisms likely affect the response of the cortex to MD. Homeostatic processes that work to preserve a certain level of cortical drive are known to operate in neuronal development (Turrigiano & Nelson, 2004) and may contribute to the ability of binocular neurons to undergo ocular dominance plasticity after deprivation (Desai, Cudmore, Nelson, & Turrigiano, 2002; Mrsic-Flogel et al., 2007). Nondeprived inputs strengthen only after deprived inputs are weakened (Frenkel & Bear, 2004), and pathways of synaptic scaling, the global (or cellwide) modulation of synapses, may be operating. TNFα, a glial secreted cytokine which acts as a positive scaling factor by increasing synaptic GluR1 and mEPSC amplitudes, is necessary for scaling up synaptic strength in vitro (Stellwagen, Beattie, Seo, & Malenka, 2005; Stellwagen & Malenka, 2006) and for the increase in amplitude of nondeprived inputs after MD (Kaneko et al., 2008). Arc, a negative scaling factor that increases AMPAR endocytosis (Chowdhury et al., 2006; Rial Verde, Lee-Osbourne, Worley, Malinow, & Cline, 2006), may also influence ocular dominance plasticity (McCurry, Tropea, Wang, & Sur, 2007).

Inhibitory networks may provide an additional circuit mechanism for modulating input strength during MD. Somatic inhibition on excitatory pyramidal cells would allow for instructive gating of precisely correlated inputs by preventing the backpropagation, as well as subsequent strengthening, of imprecisely timed inputs (Bi & Poo, 2001; Song, Miller, & Abbott, 2000; Pouille & Scanziani, 2001; Hensch, 2005). Gap junctions between parvalbumin-expressing inhibitory cells would also allow for tightly coupled inputs to drive networks of inhibitory cells more strongly and facilitate discriminative responsiveness of pyramidal cells (Galarreta & Hestrin, 2001; Hensch, 2005). Endocannabinoid signaling on presynaptic terminals of layer 2/3 are necessary for plasticity, and these synapses may modulate the drive from the supragranular inhibitory network (C. Liu, Heynen, Shuler, & Bear, 2008).

Structural Plasticity and Permissive Changes in Extracellular Matrix Increasing evidence supports the role of proteases and perineuronal nets (PNN) of extracellular matrix in regulating the ability of cortex to respond to MD. Degradation of chondroitin-sulfate proteoglycans (CSPGs) leads to the reactivation of ocular dominance plasticity in adult cortex (Pizzorusso et al., 2002, 2006). The protease tissue plasminogen activator (tPA), which cleaves extracellular matrix and other molecules, is expressed during juvenile MD and is necessary for functional plasticity in the adult (Mataga, Nagai, & Hensch, 2002; Müller & Greisinger, 1998). Application of tPA enhances spine motility, and loss of tPA prevents the loss of superficial spines after 4 days of MD (Oray et al., 2004; Mataga et al., 2004). Extracellular matrix could have a restrictive effect on ocular dominance plasticity by constraining spine motility and axonal growth or by imposing structurally mature functional elements onto intracortical inhibitory cells. Parvalbumin-expressing GABAergic cells become ensheathed in PNNs as the cortex matures (Härtig et al., 2001); degradation of PNNs may reduce the efficacy of inhibitory input by altering the ionic or chemical milieu and allow for plasticity. Mice lacking myelination factors Nogo-66 receptor and Nogo-A/B exhibit ocular dominance plasticity after brief MD in adulthood as well as a prolonged critical period (McGee, Yang, Fischer, Daw, & Strittmatter, 2005), suggesting that extracellular factors strongly constrain plasticity.

Gene Screens for Novel Plasticity Factors The use of gene microarrays to screen for differences in cortical gene expression under different conditions has facilitated the discovery of novel pathways and functional molecules involved in ocular dominance plasticity. A screen comparing the expression of normal and MD cortex at different ages revealed common and age-specific pathways modulated by MD (Madjan & Shatz, 2006). A comparison of V1 at different ages and with MD cortex showed an upregulation of actin-stabilizing genes, including the calcium sensor, cardiac troponin C, and myelinating factors, which were reversed with MD (Lyckman et al., 2008). Comparisons of dark-reared with normal V1 found a reduction in genes with a role in functional inhibition, reflecting a maturational delay,

while MD and normal V1 comparisons identified a number of growth factor and immunomodulatory factors that were upregulated in response to MD (Tropea et al., 2006).

Summary and conclusion

Retinotopic and feature selective maps constitute key organizational principles of the visual pathway. Intrinsic genetic programs and activity-dependent processes both play a role in setting up the structure and function of these maps. In addition, patterns of activity interact with programs of gene expression as they modulate signaling pathways within the cell. Understanding specific mechanisms of how visual stimulus feature maps are assembled and modified in response to experience is central to identifying fundamental processes of neural circuit development and plasticity.

REFERENCES

Angelucci, A., Clasca, F., Bricolo, E., Cramer, K. S., & Sur, M. (1997). Experimentally induced retinal projections to the ferret auditory thalamus: Development of clustered eye-specific patterns in a novel target. *J. Neurosci.*, *17*(6), 2040–2055.

Antonini, A., Fagiolini, M., & Stryker, M. P. (1999). Anatomical correlates of functional plasticity in mouse visual cortex. *J. Neurosci.*, *19*(11), 4388–4406.

Antonini, A., & Stryker, M. P. (1996). Plasticity of geniculocortical afferents following brief or prolonged monocular occlusion in the cat. *J. Comp. Neurol.*, *369*(1), 64–82.

Berardi, N., Pizzorusso, T., Ratto, G. M., & Maffei, L. (2003). Molecular basis of plasticity in the visual cortex. *Trends Neurosci.*, *26*, 369–378.

Bi, G., & Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.*, *24*, 139–166.

Bjartmar, L., Huberman, A. D., Ullian, E. M., Rentería, R. C., Liu, X., Xu, W., et al. (2006). Neuronal pentraxins mediate synaptic refinement in the developing visual system. *J. Neurosci.*, *26*(23), 6269–6281.

Bolz, J., Uziel, D., Muhlfriedel, S., Gullmar, A., Peuckert, C., Zarbalis, K., et al. (2004). Multiple roles of ephrins during the formation of thalamocortical projections: Maps and more. *J. Neurobiol.*, *59*(1), 82–94.

Bosking, W. H., Crowley, J. C., & Fitzpatrick, D. (2002). Spatial coding of position and orientation in primary visual cortex. *Nat. Neurosci.*, *5*, 874–882.

Bovolenta, P. (2005). Morphogen signaling at the vertebrate growth cone: A few cases or a general strategy? *J. Neurobiol.*, *64*(4), 405–416.

Brown, A., Yates, P. A., Burrola, P., Ortuno, D., Vaidya, A., Jessell, T. M., et al. (2000). Topographic mapping from the retina to the midbrain is controlled by relative but not absolute levels of EphA receptor signaling. *Cell*, *102*, 77–88.

Brunet, I., Weinl, C., Piper, M., Trembleau, A., Volovitch, M., Harris, W., et al. (2005). The transcription factor Engrailed-2 guides retinal axons. *Nature*, *438*(7064), 94–98.

Butts, D. A., Kanold, P. O., & Shatz, C. J. (2007). A burst-based "Hebbian" learning rule at retinogeniculate synapses links retinal waves to activity-dependent refinement. *PLoS Biol.*, *5*(3), 361.

Cang, J., Kaneko, M., Yamada, J., Woods, G., Stryker, M. P., & Feldheim, D. A. (2005). Ephrin-As guide the formation of functional maps in the visual cortex. *Neuron*, *48*, 577–589.

Cang, J., Niell, C. M., Liu, X., Pfeiffenberger, C., Feldheim, D. A., & Stryker, M. P. (2008). Selective disruption of one Cartesian axis of cortical maps and receptive fields by deficiency in ephrin-As and structured activity. *Neuron*, *57*(4), 511–523.

Carri, N. G., Bengtsson, H., Charette, M. F., & Ebendal, T. (1998). BMPR-II expression and OP-1 effects in developing chicken retinal explants. *NeuroReport*, *9*(6), 1097–1101.

Chandrasekaran, A. R., Plas, D. T., Gonzalez, E., & Crair, M. C. (2005). Evidence for an instructive role of retinal activity in retinotopic map refinement in the superior colliculus of the mouse. *J. Neurosci.*, *25*(29), 6929–6938.

Chapman, B., Stryker, M. P., & Bonhoeffer, T. (1996). Development of orientation-preference maps in ferret primary visual cortex, *J. Neurosci.*, *16*, 6443–6453.

Charron, F., & Tessier-Lavigne, M. (2005). Novel brain wiring functions for classical morphogens: A role as graded positional cues in axon guidance. *Development*, *132*(10), 2251–2262.

Chen, C., & Regehr, W. G. (2000). Developmental remodeling of the retinogeniculate synapse. *Neuron*, *28*(3), 955–966.

Chierzi, S., Ratto, G. M., Verma, P., & Fawcett, J. W. (2005). The ability of axons to regenerate their growth cones depends on axonal type and age, and is regulated by calcium, cAMP and ERK. *Eur. J. Neurosci.*, 21, 2051–2062.

Chowdhury, S., Shepherd, J. D., Okuno, H., Lyford, G., Petralia, R. S., Plath, N., et al. (2006). Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. *Neuron*, *52*, 445–459.

Colello, R. J., & Guillery, R. W. (1990). The early development of retinal ganglion cells with uncrossed axons in the mouse: Retinal position and axonal course. *Development*, *108*, 515–523.

Constantine-Paton, M., & Law, M. I. (1978). Eye-specific termination bands in tecta of three-eyed frogs. *Science*, *202*(4368), 639–641.

Coppola, D. M., & White, L. E. (2004). Visual experience promotes the isotropic representation of orientation preference. *Visual Neurosci.*, *21*, 39–51.

Cork, R. J., Namkung, Y., Shin, H. S., & Mize, R. R. (2001). Development of the visual pathway disrupted in mice with a targeted disruption of the calcium channel beta(3)-subunit gene. *J. Comp. Neurol.*, *440*(2), 177–191.

Cynader, M., Berman, N., & Hein, A. (1976). Recovery of function in cat visual cortex following prolonged deprivation. *Exp. Brain Res.*, *25*(2), 139–156.

Dehay, C., & Kennedy, H. (2007). Cell-cycle control and cortical development. *Nat. Rev. Neurosci.*, *8*(6), 438–450.

Demas, J., Sagdullaev, B. T., Green, E., Jaubert-Miazza, L., McCall, M. A., Gregg, R. G., et al. (2006). Failure to maintain eye-specific segregation in nob, a mutant with abnormally patterned retinal activity. *Neuron*, *50*(2), 247–259.

Demyanenko, G. P., & Maness, P. F. (2003). The L1 cell adhesion molecule is essential for topographic mapping of retinal axons. *J. Neurosci.*, *23*(2), 530–538.

Desai, N. S., Cudmore, R. H., Nelson, S. B., & Turrigiano, G. G. (2002). Critical periods for experience-dependent synaptic scaling in visual cortex. *Nat. Neurosci.*, *5*(8), 783–789.

Di Cristo, G., Berardi, N., Cancedda, L., Pizzorusso, T., Putignano, E., Ratto, G. M., et al. (2001). Requirement of ERK activation for visual cortical plasticity. *Science*, *292*(5525), 2337–2340.

Di Cristo, G., Chattopadhyaya, B., Kuhlman, S. J., Fu, Y., Bélanger, M. C., Wu, C. Z., et al. (2007). Activity-dependent PSA expression regulates inhibitory maturation and onset of critical period plasticity. *Nat. Neurosci.*, *10*(12), 1569–1577.

Dölen, G., Osterweil, E., Rao, B. S., Smith, G. B., Auerbach, B. D., Chattarji, S., et al. (2007). Correction of fragile X syndrome in mice. *Neuron*, *56*(6), 955–962.

Doran, N. N., & LeDoux, J. E. (1999). Organization of projections to the lateral amygdale from auditory and visual areas of the thalamus in the rat. *J. Comp. Neurol.*, *430*, 235–249.

Dräger, U. C., & Olsen, J. F. (1980). Origins of crossed and uncrossed retinal projections in pigmented and albino mice. *J. Comp. Neurol.*, *191*(3), 383–412.

Ellsworth, C. A., Lyckman, A. W., Feldheim, D. A., Flanagan, J. G., & Sur, M. (2005). Ephrin-A2 and -A5 influence patterning of normal and novel retinal projections to the thalamus: Conserved mapping mechanisms in visual and auditory thalamic targets. *J. Comp. Neurol.*, *488*, 140–151.

Erskine, L., Williams, S. E., Brose, K., Kidd, T., Rachel, R. A., Goodman, C. S., et al. (2000). Retinal ganglion cell axon guidance in the mouse optic chiasm: Expression and function of robos and slits. *J. Neurosci.*, *20*(13), 4975–4982.

Fagiolini, M., & Hensch, T. K. (2000). Inhibitory threshold for critical-period activation in primary visual cortex. *Nature*, *404*(6774), 183–186.

Fagiolini, M., Pizzorusso, T., Berardi, N., Domenici, L., & Maffei, L. (1994). Functional postnatal development of the rat primary visual cortex and the role of visual experience: Dark rearing and monocular deprivation. *Vis. Res.*, *34*(6), 709–720.

Farley, B. J., Yu, H., Jin, D. Z., & Sur, M. (2007). Alteration of visual input results in a coordinated reorganization of multiple visual cortex maps. *J. Neurosci.*, *27*(38), 10299–10310.

Feldheim, D. A., Kim, Y. I., Bergemann, A. D., Frisen, J., Barbacid, M., & Flanagan, J. G. (2000). Genetic analysis of ephrin-A2 and ephrin-A5 shows their requirement in multiple aspects of retinocollicular mapping. *Neuron*, *25*, 563–574.

Feldheim, D. A., Vanderhaeghen, P., Hansen, M. J., Frisen, J., Lu, Q., Barbacid, M., et al. (1998). Topographic guidance labels in a sensory projection to the forebrain. *Neuron*, *21*, 1303–1313.

Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.*, *23*, 441–471.

Figdor, M. C., & Stern, C. D. (1993). Segmental organization of embryonic diencephalon. *Nature*, *363*, 630–634.

Fischer, Q. S., Aleem, S., Zhou, H., & Pham, T. A. (2007). Adult visual experience promotes recovery of primary visual cortex from long-term monocular deprivation. *Learn. Memory*, *14*(9), 573–580.

Frégnac, Y., & Shulz, D. E. (1999). Activity-dependent regulation of receptive field properties of cat area 17 by supervised Hebbian learning. *J. Neurobiol.*, *41*(1), 69–82.

Frenkel, M. Y., & Bear, M. F. (2004). How monocular deprivation shifts ocular dominance in visual cortex of young mice. *Neuron*, *44*(6), 917–923.

Frisen, J., Yates, P. A., McLaughlin, T., Friedman, G. C., O'Leary, D. D., & Barbacid, M. (1998). Ephrin-A5 (AL-1/RAGS) is essential for proper retinal axon guidance and topographic mapping in the mammalian visual system. *Neuron*, *20*, 235–243.

Frost, D. O. (1982). Anomalous visual connections to somatosensory and auditory systems following brain lesions in early life. *Brain Res.*, *255*, 627–635.

Frost, D. O., & Metin, C. (1985). Induction of functional retinal projections to the somatosensory system. *Nature*, *317*, 162–164.

Fukuchi-Shimogori, T., & Grove, E. A. (2001). Neocortex patterning by the secreted signaling molecule FGF8. *Science*, *294*(5544), 1071–1074.

Galarreta, M., & Hestrin, S. (2001). Spike transmission and synchrony detection in networks of GABAergic interneurons. *Science*, *292*(5525), 2295–2299.

Gao, W. J., Wormington, A. B., Newman, D. E., & Pallas, S. L. (2000). Development of inhibitory circuitry in visual and auditory cortex of postnatal ferrets: Immunocytochemical localization of calbindin- and parvalbumin-containing neurons. *J. Comp. Neurol.*, *422*(1), 140–157.

Gianfranceschi, L., Siciliano, R., Walls, J., Morales, B., Kirkwood, A., Huang, Z. J., et al. (2003). Visual cortex is rescued from the effects of dark rearing by overexpression of BDNF. *Proc. Natl. Acad. Sci. USA*, *100*(21), 12486–12491.

Godement, P., Salaun, J., & Imbert, M. (1984). Prenatal and postnatal development of retinogeniculate and retinocollicular projections in the mouse. *J. Comp. Neurol.*, *230*, 552–575.

Gomez, L. L., Alam, S., Smith, K. E., Horne, E., & Dell'Acqua, M. L. (2002). Regulation of A-kinase anchoring protein 79/150-cAMP-dependent protein kinase postsynaptic targeting by NMDA receptor activation of calcineurin and remodeling of dendritic actin. *J. Neurosci.*, *22*, 7027–7044.

Gordon, J. A., & Stryker, M. P. (1996). Experience-dependent plasticity of binocular responses in the primary visual cortex of the mouse. *J. Neurosci.*, *16*(10), 3274–3286.

Grove, E. A., & Fukuchi-Shimogori, T. (2003). Generating the cerebral cortical area map. *Annu. Rev. Neurosci.*, *26*, 355–380.

Grubb, M. S., Rossi, F. M., Changeux, J. P., & Thompson, I. D. (2003). Abnormal functional organization in the dorsal lateral geniculate nucleus of mice lacking the beta 2 subunit of the nicotinic acetylcholine receptor. *Neuron*, *40*(6), 1161–1172.

Guido, W. (2008). Refinement of the retinogeniculate pathway. *J. Physiol*, *586*, 4357–4362.

Gurung, B., & Fritzsch, B. (2004). Time course of embryonic midbrain and thalamic auditory connection development in mice as revealed by carbocyanine dye tracing. *J. Comp. Neurol.*, *479*(3), 309–327.

Hanover, J. L., Huang, Z. J., Tonegawa, S., & Stryker, M. P. (1999). Brain-derived neurotrophic factor overexpression induces precocious critical period in mouse visual cortex. *J. Neurosci.*, *19*(22), RC40.

Hansen, M. J., Dallal, G. E., & Flanagan, J. G. (2004). Retinal axon response to ephrin-As shows a graded, concentration-dependent transition from growth promotion to inhibition. *Neuron*, *42*(5), 717–730.

Harris, W. A. (1980). The effects of eliminating impulse activity on the development of the retinotectal projection in salamanders. *J. Comp. Neurol.*, *194*, 303–317.

Härtig, W., Singer, A., Grosche, J., Brauer, K., Ottersen, O. P., & Brückner, G. (2001). Perineuronal nets in the rat

medial nucleus of the trapezoid body surround neurons immunoreactive for various amino acids, calcium-binding proteins and the potassium channel subunit Kv3.1b. *Brain Res.*, *899*(1–2), 123–133.

He, H. Y., Hodos, W., & Quinlan, E. M. (2006). Visual deprivation reactivates rapid ocular dominance plasticity in adult visual cortex. *J. Neurosci.*, *26*(11), 2951–2955.

Hebb, D. O. (1949). *The organization of behavior*. New York: John Wiley and Sons.

Hehr, C. L., Hocking, J. C., & McFarlane, S. (2005). Matrix metalloproteinases are required for retinal ganglion cell axon guidance at select decision points. *Development*, *132*(15), 3371–3379.

Heldt, S., Sudin, V., Willott, J. F., & Falls, W. A. (2000). Posttraining lesions of the amygdale interfere with fear-potentiated startle to both visual and auditory conditioned stimuli in C57BL/6J mice. *Behav. Neurosci.*, *114*, 749–759.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nat. Rev. Neurosci.*, *6*(11), 877–888.

Hensch, T. K., Fagiolini, M., Mataga, N., Stryker, M. P., Baekkeskov, S., & Kash, S. F. (1998). Local GABA circuit control of experience-dependent plasticity in developing visual cortex. *Science*, *282*(5393), 1504–1508.

Herrera, E., Brown, L., Aruga, J., Rachel, R. A., Dölen, G., Mikoshiba, K., et al. (2003). Zic2 patterns binocular vision by specifying the uncrossed retinal projection. *Cell*, *114*(5), 545–557.

Heynen, A. J., Yoon, B. J., Liu, C. H., Chung, H. J., Huganir, R. L., & Bear, M. F. (2003). Molecular mechanism for loss of visual cortical responsiveness following brief monocular deprivation. *Nat. Neurosci.*, *6*(8), 854–862.

Hindges, R., McLaughlin, T., Genoud, N., Henkemeyer, M., & O'Leary, D. D. (2002). EphB forward signaling controls directional branch extension and arborization required for dorsal-ventral retinotopic mapping. *Neuron*, *35*(3), 475–487.

Hofer, S. B., Mrsic-Flogel, T. D., Bonhoeffer, T., & Hübener, M. (2006). Prior experience enhances plasticity in adult visual cortex. *Nat. Neurosci.*, *9*(1), 127–132.

Hopker, V. H., Shewan, D., Tessier-Lavigne, M., Poo, M., & Holt, C. (1999). Growth-cone attraction to netrin-1 is converted to repulsion by laminin-1. *Nature*, *401*(6748), 69–73.

Huang, Z. J., Kirkwood, A., Pizzorusso, T., Porciatti, V., Morales, B., Bear, M. F., et al. (1999). BDNF regulates the maturation of inhibition and the critical period of plasticity in mouse visual cortex. *Cell*, *98*(6), 739–755.

Hubel, D. H., & Wiesel, T. N. (1963). Shape and arrangement of columns in cat's striate cortex. *J. Physiol.*, *165*, 559–568.

Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J. Physiol.*, *206*, 419–436.

Huberman, A. D., Murray, K. D., Warland, D. K., Feldheim, D. A., & Chapman, B. (2005). Ephrin-As mediate targeting of eye-specific projections to the lateral geniculate nucleus. *Nat. Neurosci.*, *8*(8), 1013–1021.

Huh, G. S., Boulanger, L. M., Du, H., Riquelme, P. A., Brotz, T. M., & Shatz, C. J. (2000). Functional requirement for class I MHC in CNS development and plasticity. *Science*, *290*(5499), 2155–2159.

Inatani, M. (2005). Molecular mechanisms of optic axon guidance. *Naturwissenschaften*, *92*(12), 549–561.

Inoue, T., Nakamura, S., & Osumi, N. (2000). Fate mapping of the mouse prosencephalic neural plate. *Dev. Biol. 219*(2), 373–383.

Jaubert-Miazza, L., Green, E., Lo, F. S., Bui, K., Mills, J., & Guido, W. (2005). Structural and functional composition of the developing retinogeniculate pathway in the mouse. *Visual Neurosci.*, *22*(5), 661–676.

Jiang, B., Treviño, M., & Kirkwood, A. (2007). Sequential development of long-term potentiation and depression in different layers of the mouse visual cortex. *J. Neurosci.*, *27*(36), 9648–9652.

Job, C., & Tan, S. (2003). Constructing the mammalian neocortex: The role of intrinsic factors. *Dev. Biol.*, *257*, 221–232.

Jones, E. G. (1985). *The thalamus*. New York: Plenum.

Jones, E. G., & Rubenstein, J. L. R. (2004). Expression of regulatory genes during differentiation of thalamic nuclei in mouse and monkey. *J. Comp. Neurol.*, *47*, 55–80.

Kalil, R. E., & Schneider, G. E. (1975). Abnormal synaptic connections of the optic tract in the thalamus after midbrain lesions in newborn hamsters. *Brain Res*, *100*, 690–698.

Kaneko, M., Stellwagen, D., Malenka, R. C., & Stryker, M. P. (2008). Tumor necrosis factor–alpha mediates one component of competitive, experience-dependent plasticity in developing visual cortex. *Neuron*, *58*(5), 673–680.

Katz, L. C., & Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. *Science*, *274*, 1133–1138.

Kolpak, A., Zhang, J., & Bao, Z. Z. (2005). Sonic hedgehog has a dual effect on the growth of retinal ganglion axons depending on its concentration. *J. Neurosci.*, *25*(13), 3432–3441.

Leamey, C. A., Glendining, K. A., Kreiman, G., Kang, N. D., Wang, K. H., Fassler, R., et al. (2008). Differential gene expression between sensory neocortical areas: Potential roles for ten_m3 and Bcl6 in patterning visual and somatosensory pathways. *Cereb. Cortex*, *18*(1), 53–66.

Leamey, C. A., Merlin, S., Lattouf, P., Sawatari, A., Zhou, X., Demel, N., et al. (2007). Ten_m3 regulates eye-specific patterning in the mammalian visual pathway and is required for binocular vision. *PLoS Biol.*, *5*, e241.

Lee, R., Petros, T. J., & Mason, C. A. (2008). Zic2 regulates retinal ganglion cell axon avoidance of ephrinB2 through inducing expression of the guidance receptor EphB1. *J. Neurosci.*, *28*(23), 5910–5919.

Lee, W. C., Huang, H., Feng, G., Sanes, J. R., Brown, E. N., So, P. T., et al. (2006). Dynamic remodeling of dendritic arbors in GABAergic interneurons of adult visual cortex. *PLoS Biol.*, *4*(2), e29.

Li, Y., Fitzpatrick, D., & White, L. E. (2006). The development of direction selectivity in ferret visual cortex requires early visual experience. *Nat. Neurosci.*, *9*, 676–681.

Liu, C. H., Heynen, A. J., Shuler, M. G., & Bear, M. F. (2008). Cannabinoid receptor blockade reveals parallel plasticity mechanisms in different layers of mouse visual cortex. *Neuron*, *58*, 340–345.

Liu, X., & Chen, C. (2008). Different roles for AMPA and NMDA receptors in transmission at the immature retinogeniculate synapse. *J. Neurophysiol.*, *99*(2), 629–643.

Luo, L., & Flanagan, J. G. (2007). Development of continuous and discrete neural maps. *Neuron*, *56*(2), 284–300.

Lyckman, A. W., Horng, S., Leamey, C. A., Tropea, D., Watakabe, A., Van Wart, A., et al. (2008). Gene expression patterns in visual cortex during the critical period: Synaptic stabilization and reversal by visual deprivation. *Proc. Natl. Acad. Sci. USA*, *105*(27), 9409–9414.

Lyckman, A. W., Jhaveri, S., Feldheim, D. A., Vanderhaeghen, P., Flanagan, J. G., & Sur, M. (2001). Enhanced plasticity of

retinothalamic projections in an ephrin-A2/A5 double mutant. *J. Neurosci.*, *21*, 7684–7690.

Majdan, M., & Shatz, C. J. (2006). Effects of visual experience on activity-dependent gene regulation in cortex. *Nat. Neurosci.*, *9*, 650–659.

Majewska, A., & Sur, M. (2003). Motility of dendritic spines in visual cortex in vivo: Changes during the critical period and effects of visual deprivation. *Proc. Natl. Acad. Sci. USA*, *100*(26), 16024–16029.

Malinow, R., & Malenka, R. C. (2002). AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.*, *25*, 103–126.

Mann, F., Harris, W. A., & Holt, C. E. (2004). New views on retinal axon development: A navigation guide. *Int. J. Dev. Biol.*, *48*(8–9), 957–964.

Marino, J., Schummers, J., Lyon, D. C., Schwabe, L., Beck, O., Wiesing, P., et al. (2005). Invariant computations in local cortical networks with balanced excitation and inhibition. *Nat. Neurosci*, *8*, 194–201.

Mataga, N., Mizuguchi, Y., & Hensch, T. K. (2004). Experience-dependent pruning of dendritic spines in visual cortex by tissue plasminogen activator. *Neuron*, *44*(6), 1031–1041.

Mataga, N., Nagai, N., & Hensch, T. K. (2002). Permissive proteolytic activity for visual cortical plasticity. *Proc. Natl. Acad. Sci. USA*, *99*(11), 7717–7721.

McCurry, C., Tropea, D., Wang, K. H., & Sur, M. (2007). *A role for Arc in constraining ocular dominance plasticity in adult visual cortex.* Program No. 1304/B22 Neuroscience Meeting Planner. San Diego: Society for Neuroscience.

McGee, A. W., Yang, Y., & Fischer, Q. S., Daw, N. W., & Strittmatter, S. M. (2005). Experience-driven plasticity of visual cortex limited by myelin and Nogo receptor. *Science*, *309*(5744), 2222–2226.

McLaughlin, T., Hindges, R., Yates, P. A., & O'Leary, D. D. (2003). Bifunctional action of ephrin-B1 as a repellent and attractant to control bidirectional branch extension in dorsal-ventral retinotopic mapping. *Development*, *130*(11), 2407–2418.

McLaughlin, T., Torborg, C. L., Feller, M. B., O'Leary, D. D. (2003). Retinotopic map refinement requires spontaneous retinal waves during a brief critical period of development. *Neuron*, *40*(6), 1147–1160.

Meister, M., Wong, R. O., Baylor, D. A., & Shatz, C. J. (1991). Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science*, *252*(5008), 939–943.

Meliza, C. D., & Dan, Y. (2006). Receptive-field modification in rat visual cortex induced by paired visual stimulation and single-cell spiking. *Neuron*, *49*(2), 183–189.

Monnier, P. P., Sierra, A., Macchi, P., Deitinghoff, L., Andersen, J. S., Mann, M., et al. (2002). RGM is a repulsive guidance molecule for retinal axons. *Nature*, *419*(6905), 392–395.

Mower, A. F., Liao, D. S., Nestler, E. J., Neve, R. L., & Ramoa, A. S. (2002). cAMP/Ca2+ response element-binding protein function is essential for ocular dominance plasticity. *J. Neurosci.*, *22*(6), 2237–2245.

Mower, G. D. (1991). The effect of dark rearing on the time course of the critical period in cat visual cortex. *Brain Res. Dev. Brain Res.*, *58*(2), 151–158.

Mrsic-Flogel, T. D., Hofer, S. B., Ohki, K., Reid, R. C., Bonhoeffer, T., & Hübener, M. (2007). Homeostatic regulation of eye-specific responses in visual cortex during ocular dominance plasticity. *Neuron*, *54*(6), 961–972.

Müller, C. M., & Griesinger, C. B. (1998). Tissue plasminogen activator mediates reverse occlusion plasticity in visual cortex. *Nat. Neurosci.*, *1*(1), 47–53.

Nakagawa, Y., & O'Leary, D. D. (2001). Combinatorial expression patterns of LIM-homeodomain and other regulatory genes parcellate developing thalamus. *J. Neurosci.*, *21*(8), 2711–2725.

Nakagawa, Y., & O'Leary, D. D. (2002). Patterning centers, regulatory genes and extrinsic mechanisms controlling arealization of the neocortex. *Curr. Opin. Neurobiol.*, *12*(1), 14–25.

Nakamoto, M., Cheng, H. J., Friedman, G. C., McLaughlin, T., Hansen, M. J., Yoon, C. H., et al. (1996). Topographically specific effects of ELF-1 on retinal axon guidance in vitro and retinal axon mapping in vivo. *Cell*, *86*(5), 755–766.

Newton, J. R., Ellsworth, C., Miyakawa, T., Tonegawa, S., & Sur, M. (2004). Acceleration of visually cued conditioned fear through the auditory pathway. *Nat. Neurosci.*, *7*(9), 968–973.

Nicol, X., Voyatzis, S., Muzerelle, A., Narboux-Nême, N., Südhof, T. C., Miles, R., et al. (2007). cAMP oscillations and retinal activity are permissive for ephrin signaling during the establishment of the retinotopic map. *Nat. Neurosci.*, *10*(3), 340–347.

O'Leary, D. D. (1989). Do cortical areas emerge from a protocortex? *Trends Neurosci.*, *12*(10), 400–406.

Oray, S., Majewska, A., & Sur, M. (2004). Dendritic spine dynamics are regulated by monocular deprivation and extracellular matrix degradation. Neuron, *44*(6), 1021–1030.

Oster, S. F., Bodeker, M. O., He, F., & Sretavan, D. W. (2003). Invariant Sema5A inhibition serves an ensheathing function during optic nerve development. *Development*, *130*(4), 775–784.

Pak, W., Hindges, R., Lim, Y. S., Pfaff, S. L., & O'Leary, D. D. (2004). Magnitude of binocular vision controlled by islet-2 repression of a genetic program that specifies laterality of retinal axon pathfinding. Cell, *119*, 567–578.

Pallas, S. L., Hahm, J., & Sur, M. (1994). Morphology of retinal axons induced to arborize in a novel target, the medial geniculate nucleus. I. Comparison with arbors in normal targets. *J. Comp. Neurol.*, *349*(3), 343–362.

Pallas, S. L., Roe, A. W., & Sur, M. (1990). Visual projections induced into the auditory pathway of ferrets. I. Novel inputs to primary auditory cortex (AI) from the LP/pulvinar complex and the topography of the MGN-AI projection. *J. Comp. Neurol.*, *298*(1), 50–68.

Pfeiffenberger, C., Cutforth, T., Woods, G., Yamada, J., Renteria, R. C., & Copenhagen, D. R. (2005). Ephrin-As and neural activity are required for eye-specific patterning during retinogeniculate mapping. *Nat. Neurosci.*, *8*, 1022–1027.

Pfeiffenberger, C., Yamada, J., & Feldheim, D. A. (2006). Ephrin-As and patterned retinal activity act together in the development of topographic maps in the primary visual system. *J. Neurosci.*, *26*, 12873–12884.

Pham, T. A., Rubenstein, J. L., Silva, A. J., Storm, D. R., & Stryker, M. P. (2001). The CRE/CREB pathway is transiently expressed in thalamic circuit development and contributes to refinement of retinogeniculate axons. *Neuron*, *31*(3), 409–420.

Pizzorusso, T., Medini, P., Berardi, N., Chierzi, S., Fawcett, J. W., & Maffei, L. (2002). Reactivation of ocular dominance plasticity in the adult visual cortex. *Science*, *298*, 1248–1251.

Pizzorusso, T., Medini, P., Landi, S., Baldini, S., Berardi, N., & Maffei, L. (2006). Structural and functional recovery from early monocular deprivation in adult rats. *Proc. Natl. Acad. Sci. USA*, *103*(22), 8517–8522.

Plump, A. S., Erskine, L., Sabatier, C., Brose, K., Epstein, C. J., Goodman, C. S., et al. (2002). Slit1 and Slit2 cooperate to prevent premature midline crossing of retinal axons in the mouse visual system. *Neuron*, *33*(2), 219–232.

Pouille, F., & Scanziani, M. (2001). Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science*, *293*(5532), 1159–1163.

Ragsdale, C. W., & Grove, E. A. (2001). Patterning the mammalian cerebral cortex. *Curr. Opin. Neurobiol.*, *11*(1), 50–58.

Rakic, P. (1988). Specification of cerebral cortical areas. *Science*, *242*, 170–176.

Rao, S. C., Toth, L. J., & Sur, M. (1997). Optically imaged maps of orientation preference in primary visual cortex of cats and ferrets. *J. Comp. Neurol.*, *387*(3), 358–370.

Rial Verde, E. M., Lee-Osbourne, J., Worley, P. F., Malinow, R., & Cline, H. T. (2006). Increased expression of the immediate-early gene arc/arg3.1 reduces AMPA receptormediated synaptic transmission. *Neuron*, *52*, 461–474.

Ringstedt, T., Braisted, J. E., Brose, K., Kidd, T., Goodman, C., & Tessier-Lavigne, M., et al. (2000). Slit inhibition of retinal axon growth and its role in retinal axon pathfinding and innervation patterns in the diencephalon. *J. Neurosci.*, *20*(13), 4983–4991.

Rodriguez, J., Esteve, P., Weinl, C., Ruiz, J. M., Fermin, Y., & Trousse, F. (2005). SFRP1 regulates the growth of retinal ganglion cell axons through the Fz2 receptor. *Nat. Neurosci.*, *8*(10), 1301–1309.

Roe, A. W., Garraghty, P. E., Esguerra, M., & Sur, M. (1993). Experimentally induced visual projections to the auditory thalamus in ferrets: Evidence for a W cell pathway. *J. Comp. Neurol.*, *334*(2), 263–280.

Roe, A. W., Hahm, J. O., & Sur, M. (1991). Experimentally induced establishment of visual topography in auditory thalamus. *Soc. Neurosci. Abstracts*, *17*, 898.

Roe, A. W., Pallas, S. L., Hahm, J. O., & Sur, M. (1990). A map of visual space induced in primary auditory cortex. *Science*, *250*(4982), 818–820.

Roe, A. W., Pallas, S. L., Kwon, Y. H., & Sur, M. (1992). Visual projections routed to the auditory pathway in ferrets: Receptive fields of visual neurons in primary auditory cortex. *J. Neurosci.*, *12*(9), 3651–3664.

Rogan, M. T., & LeDoux, J. E. (1995). LTP is accompanied by commensurate enhancement of auditor-evoked responses in a fear conditioning circuit. *Neuron*, *15*, 127–136.

Rossi, F. M., Pizzorusso, T., Porciatti, V., Marubio, L. M., Maffei, L., & Changeux, J. P. (2001). Requirement of the nicotinic acetylcholine receptor beta 2 subunit for the anatomical and functional development of the visual system. *Proc. Natl. Acad. Sci. USA*, *98*(11), 6453–6458.

Rubenstein, J. L., Martinez, S., Shimamura, K., & Puelles, L. (1994). The embryonic vertebrate forebrain: The prosomeric model. *Science*, *266*(5185), 578–580.

Rubenstein, J. L., Shimamura, K., Martinez, S., & Puelles, L. (1998). Regionalization of the prosencephalic neural plate. *Annu. Rev. Neurosci.*, *21*, 445–477.

Sawtell, N. B., Frenkel, M. Y., Philpot, B. D., Nakazawa, K., Tonegawa, S., & Bear, M. F. (2003). NMDA receptor-dependent ocular dominance plasticity in adult visual cortex. *Neuron*, *38*(6), 977–985.

Schmidt, J. T., & Eisele, L. E. (1985). Stroboscopic illumination and dark rearing block the sharpening of the regenerated retinotectal map in goldfish. *Neuroscience*, *14*(2), 535–546.

Schmitt, A. M., Shi, J., Wolf, A. M., Lu, C. C., King, L. A., & Zou, Y. (2006). Wnt-Ryk signalling mediates medial-lateral retinotectal topographic mapping. *Nature*, *439*(7072), 31–37.

Schneider, G. E. (1973). Early lesions of superior colliculus: Factors affecting the formation of abnormal retinal projections. *Brain Behav. Evol.*, *8*, 73–109.

Sharma, J., Angelucci, A., & Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature*, *404*, 841–847.

Shatz, C. J. (1983). The prenatal development of the cat's retinogeniculate pathway. *J. Neurosci.*, *3*, 482–499.

Shatz, C. J., & Stryker, M. P. (1988). Prenatal tetrodotoxin infusion blocks segregation of retinogeniculate afferents. *Science*, *242*, 87–89.

Shewan, D., Dwivedy, A., Anderson, R., & Holt, C. E. (2002). Age-related changes underlie switch in netrin-1 responsiveness as growth cones advance along visual pathway. *Nat. Neurosci.*, *5*(10), 955–962.

Shimogori, T., Banuchi, V., Ng, H. Y., Strauss, J. B., & Grove, E. A. (2004). Embryonic signaling centers expressing BMP, WNT and FGF proteins interact to pattern the cerebral cortex. *Development*, *131*(22), 5639–5647.

Smith, S. L., & Trachtenberg, J. T. (2007). Experience-dependent binocular competition in the visual cortex begins at eye opening. *Nat. Neurosci.*, *10*(3), 370–375.

Somers, D. C., Nelson, S. B., & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.*, *15*, 5448–5465.

Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.*, *3*(9), 919–926.

Sperry, R. W. (1963). Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc. Natl. Acad. Sci. USA*, *50*, 703–710.

Stellwagen, D., Beattie, E. C., Seo, J. Y., & Malenka, R. C. (2005). Differential regulation of AMPA receptor and GABA receptor trafficking by tumor necrosis factor-alpha. *J. Neurosci.*, *25*(12), 3219–3228.

Stellwagen, D., & Malenka, R. C. (2006). Synaptic scaling mediated by glial TNF-alpha. *Nature*, *440*(7087), 1054–1059.

Stevens, B., Allen, N. J., Vazquez, L. E., Howell, G. R., Christopherson, K. S., Nouri, N., et al. (2007). The classical complement cascade mediates CNS synapse elimination. *Cell*, *131*(6), 1164–1178.

Stryker, M. P., & Harris, W. A. (1986). Binocular impulse blockade prevents the formation of ocular dominance columns in cat visual cortex. *J. Neurosci.*, *6*(8), 2117–2133.

Sur, M., Garraghty, P. E. & Roe, A. W. (1988). Experimentally induced visual projections into auditory thalamus and cortex. *Science*, *242*, 1437–1441.

Sur, M., & Leamey, C. A. (2001). Development and plasticity of cortical areas and networks. *Nat. Rev. Neurosci.*, *2*(4), 251–262.

Sur, M., Pallas, S. L., & Roe, A. W. (1990). Cross-modal plasticity in cortical development: Differentiation and specification of sensory neocortex. *Trends Neurosci.*, *13*(6), 227–233.

Sur, M., & Rubenstein, J. L. R. (2005). Patterning and plasticity of the cerebral cortex. *Science*, *310*, 805–810.

Suzuki, S., al-Noori, S., Butt, S. A., & Pham, T. A. (2004). Regulation of the CREB signaling cascade in the visual cortex by visual experience and neuronal activity. *J. Comp. Neurol.*, *479*, 70–83.

Swindale, N. V., Shoham, D., Grinvald, A., Bonhoeffer, T., & Hübener, M. (2000). Visual cortex maps are optimized for uniform coverage. *Nat. Neurosci.*, *3*(8), 822–826.

Taha, S., & Stryker, M. P. (2002). Rapid ocular dominance plasticity requires cortical but not geniculate protein synthesis. *Neuron*, *34*(3), 425–436.

Taha, S. A., & Stryker, M. P. (2005). Ocular dominance plasticity is stably maintained in the absence of alpha calcium calmodulin kinase II (alphaCaMKII) autophosphorylation. *Proc. Natl. Acad. Sci. USA*, *102*(45), 16438–16442.

Tavazoie, S. F., & Reid, R. C. (2000). Diverse receptive fields in the lateral geniculate nucleus during thalamocortical development. *Nat. Neurosci.*, *3*(6), 608–616.

Trachtenberg, J. T., & Stryker, M. P. (2001). Rapid anatomical plasticity of horizontal connections in the developing visual cortex. *J. Neurosci.*, *21*(10), 3476–3482.

Trachtenberg, J. T., Trepel, C., & Stryker, M. P. (2000). Rapid extragranular plasticity in the absence of thalamocortical plasticity in the developing primary visual cortex. *Science*, *287*(5460), 2029–2032.

Tropea, D., Kreiman, G., Lyckman, A., Mukherjee, S., Yu, H., Horng, S., et al. (2006). Gene expression changes and molecular pathways mediating activity-dependent plasticity in visual cortex. *Nat. Neurosci.*, *9*, 660–668.

Trousse, F., Marti, E., Gruss, P., Torres, M., & Bovolenta, P. (2001). Control of retinal ganglion cell axon growth: A new role for sonic hedgehog. *Development*, *128*(20), 3927–3936.

Turrigiano, G. G., & Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nat. Rev. Neurosci.*, *5*, 97–107.

Tuttle, R., Braisted, J. E., Richards, L. J., & O'Leary, D. D. (1998). Retinal axon guidance by region-specific cues in diencephalon. *Development*, *125*(5), 791–801.

von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, *404*(6780), 871–876.

Wang, K. H., Majewska, A., Schummers, J., Farley, B., Hu, C., Sur, M., et al. (2006). In vivo two-photon imaging reveals a role of arc in enhancing orientation specificity in visual cortex. *Cell*, *126*(2), 389–402.

Webber, C. A., Hyakutake, M. T., & McFarlane, S. (2003). Fibroblast growth factors redirect retinal axons in vitro and in vivo. *Dev. Biol.*, *263*(1), 24–34.

White, L. E., Coppola, D. M., & Fitzpatrick, D. (2001). The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature*, *411*, 1049–1052.

White, L. E., & Fitzpatrick, D. (2007). Vision and cortical map development. *Neuron*, *56*(2), 327–338.

Williams, S. E., Mann, F., Erskine, L., Sakurai, T., Wei, S., Rossi, D. J., et al. (2003). Ephrin-B2 and EphB1 mediate retinal axon divergence at the optic chiasm. *Neuron*, *39*(6), 919–935.

Wong, R. O., Meister, M., & Shatz, C. J. (1993). Transient period of correlated bursting activity during development of the mammalian retina. *Neuron*, *11*(5), 923–938.

Yang, Y., Fischer, Q. S., Zhang, Y., Baumgärtel, K., Mansuy, I. M., & Daw, N. W. (2005). Reversible blockade of experience-dependent plasticity by calcineurin in mouse visual cortex. *Nat. Neurosci.*, *8*(6), 791–796.

Yu, H., Farley, B. J., Jin, D. Z., & Sur, M. (2005). The coordinated mapping of visual space and response features in visual cortex. *Neuron*, *47*(2), 267–280.

Zhang, L. I., & Poo, M. M. (2001). Electrical activity and development of neural circuits. *Nat. Neurosci.*, *4*, Suppl., 1207–1214.

Zhou, X. H., Brandau, O., Feng, K., Oohashi, T., Ninomiya, Y., Rauch, U., et al. (2003). The murine Ten-m/Odz genes show distinct but overlapping expression patterns during development and in adult brain. *Gene Expr. Patterns*, *3*, 397–405.

Ziburkus, J., Lo, F. S., & Guido, W. (2003). Nature of inhibitory postsynaptic activity in developing relay cells of the lateral geniculate nucleus. *J. Neurophysiol.*, *90*(2), 1063–1070.

7 Synaptic Plasticity and Spatial Representations in the Hippocampus

JONATHAN R. WHITLOCK AND EDVARD I. MOSER

ABSTRACT How does the brain acquire and remember new experiences? It is believed that synaptic plasticity, the process by which synaptic connections are strengthened or weakened, is a key mechanism for information storage in the central nervous system. Long-term potentiation (LTP), the long-lasting enhancement of excitatory synaptic transmission, and long-term depression (LTD), the persistent depression of synaptic responsiveness, are experimental models of synaptic plasticity thought to reveal how synapses are modified during learning. In this chapter we focus on the properties of LTP and LTD that make them attractive functional models for memory and review key findings from studies that demonstrate a link between LTP and behavior. We then discuss how synaptic modifications can affect spatial representations expressed by hippocampal place cells, which have been used as tools for understanding how synaptic changes are implemented in neural networks. We conclude the chapter by discussing how attractor states in neural networks can aid in the storage and recall of many representations involving more than just space, and how LTP may help fine-tune shifts between attractor states during behavior.

Synaptic modifications as a means for memory: The realization of an idea

The idea that memory traces are stored as changes in synaptic efficacy is anything but new. Since the late 19th century, when the Spanish neuroanatomist Santiago Ramón y Cajal first observed spinelike structures lining the dendrites of cortical pyramidal cells, the idea has existed that the connections between nerve cells provide an anatomical substrate for memory. This idea was formalized by Donald Hebb in his 1949 work *The Organization of Behavior*, in which he formulated his famous postulate that is still one of the most quoted phrases in neuroscience: "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" (Hebb, 1949). Hebb recognized that there must be a mechanism by which a postsynaptic neuron (cell "B") can stabilize its connections with a nearby presynaptic neuron (cell "A") when it activates the postsynaptic cell strongly enough.

It was not until approximately 20 years after Hebb's seminal work that the first efforts were made that would successfully demonstrate the physical reality of long-lasting, activity-dependent modifications at synapses. Prior to the discovery of *long-term potentiation* (LTP), researchers had sought and failed to elicit long-term synaptic modifications in spinal pathways, where the observed enhancements were very short-lived (Eccles & McIntyre, 1953), and in the neocortex, whose extensively intricate neuroanatomy proved too complex to isolate responses from single synapses. It was in the hippocampus, whose straightforward laminar architecture made easy the study of monosynaptic responses (figure 7.1*A*), where LTP was discovered by Terje Lømo and Tim Bliss (Bliss & Lømo, 1973). The initial characterization was made in the dentate gyrus, the first of the three major subfields in the trisynaptic circuit of the hippocampus (figure 7.1*A*). Bliss and Lømo used a stimulating electrode to deliver brief pulses of minute electrical current to the *perforant path* (PP, figure 7.1*A*), the largest direct input from the neocortex to the hippocampus. A recording electrode was placed in the dentate gyrus (DG), the subregion of the hippocampus that receives the largest perforant path input, to record *field excitatory postsynaptic potentials* (fEPSPs) evoked in response to electrical stimulation. A fEPSP is a transient voltage deflection recorded at the tip of an extracellular recording electrode when ions flow into or out of the dendrites of large populations of cells (see traces at top of figure 7.1*B*; the responses are negative-going in this case because positive current is flowing away from the electrode). It was found that the amplitude of dentate fEPSPs, taken as a measure of synaptic strength, showed substantial increases lasting for several hours in response to brief (10-second) episodes of tetanic (15 Hz) stimulation applied to the perforant path, and that the enhancements were only expressed in the pathways that received the tetanus (figure 7.1*B*). The fact that a brief stimulus could induce changes that were (1) long lasting and (2) input specific in a structure that was known to be involved in memory formation (Scoville & Milner, 1957) immediately

JONATHAN R. WHITLOCK and EDVARD I. MOSER Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, Norwegian University of Science and Technology, Trondheim, Norway

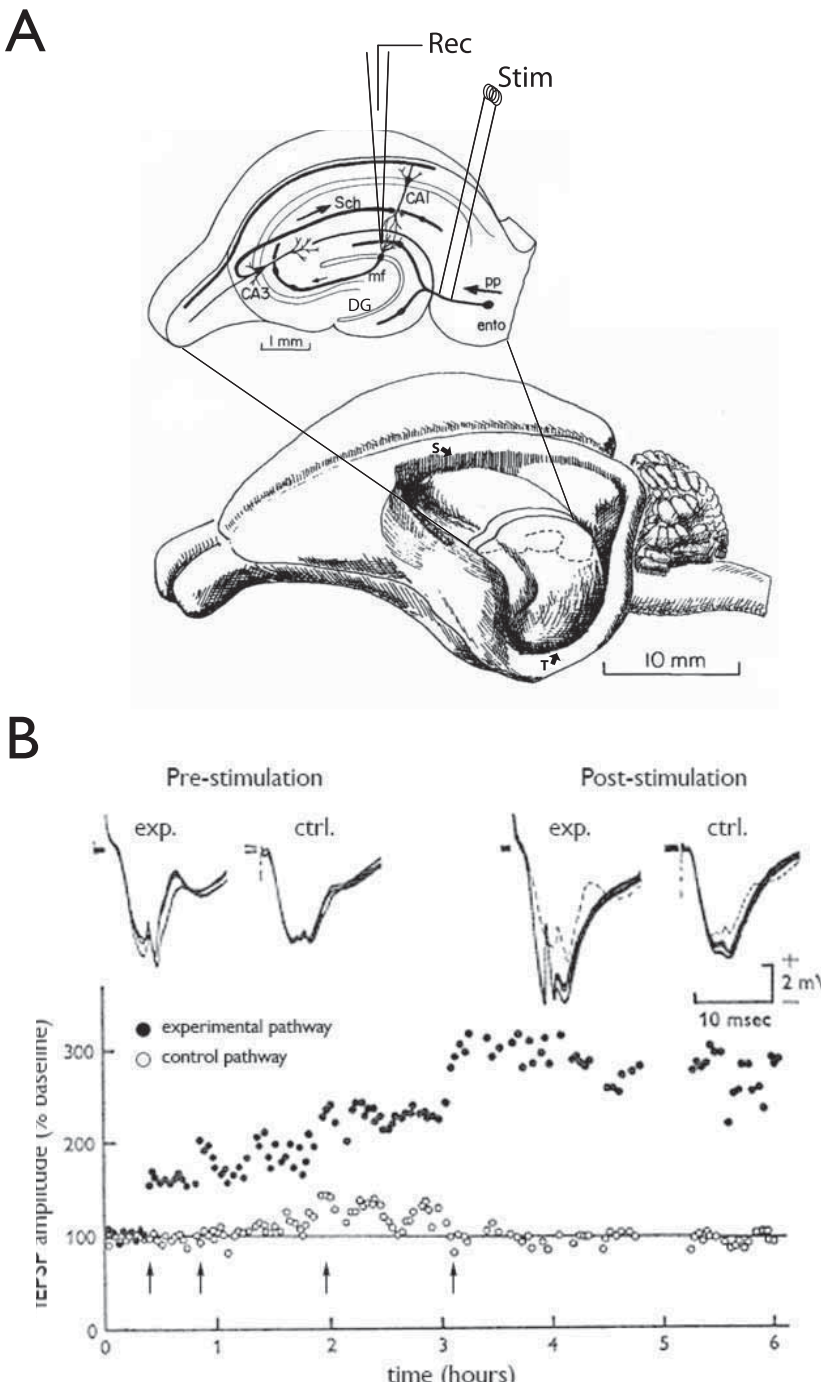


FIGURE 7.1 Anatomy of the hippocampal formation and a demonstration of hippocampal LTP. (*A*) A drawing of the rabbit brain (below) demonstrating the location of the hippocampus, oriented so that the anterior is facing left, and posterior is facing right. The long axis of the hippocampus extends from the septum (S) to the temporal cortex (T). Shown above is an enlarged cross section of the hippocampus detailing the trisynaptic connectivity between the different subregions and the placement of stimulating (Stim) and recording (Rec) electrodes by Bliss and Lømo. Abbreviations: ento, entorhinal cortex; pp, perforant path; DG, dentate gyrus; mf, mossy fiber pathway; CA3, CA1, cornu ammonis fields of the hippocampus; Sch, Schaffer collateral pathway. (Modified with kind permission of Springer Science+Business Media and from P. Andersen et al., Lamellar organization of hippocampal excitatory pathways, *Experimental Brain Research*, *13*, 222–238, © 1971.) (*B*) Example of an LTP experiment in the anaesthetized rabbit in which fEPSP amplitude served as the measure of synaptic strength (traces shown at top). The responses were obtained (left) before conditioning and (right) 2.5 hr after the 4th simulating train. On the graph (below), fEPSP response amplitude is plotted on the *y*-axis; values are expressed as a percentage of the prestimulation baseline; enhancements were specific to the pathway that received the conditioning stimulation ("experimental pathway," black dots). Arrows indicate time of tetanization. (Modified with permission from T. Bliss & T. Lømo, Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, *232*[2], 331–356, © 1973, Blackwell Publishing.)

suggested LTP as a possible model for understanding memory at the neural level.

Some years after the discovery of LTP, other labs managed to reproduce the effect, and LTP was reliably induced in hippocampal slices (Andersen, Sundberg, Sveen, & Wigstrom, 1977), paving the way for detailed studies of the underlying cellular mechanisms. Long-term potentiation was found to exist in a number of forms and in a number of brain areas, and the complementary phenomenon of *long-term depression* (LTD) was discovered. Input-specific LTD was first characterized in area CA1 of the hippocampus, where it was found that longer (15-minute) periods of more modest stimulation (1 Hz) produced a long-lasting, nonpathological decrease in the slope of fEPSPs (Dudek & Bear, 1992; Mulkey & Malenka, 1992). The possible involvement of LTP and LTD in memory formation has made *synaptic plasticity* one of the most studied phenomena in the brain, with thousands of studies over the years aimed at elucidating the cellular mechanisms for the induction, expression, and maintenance of LTP and LTD.

LTP and LTD: A short overview of cellular mechanisms

Because of its anatomical simplicity and accessibility, the vast majority of studies exploring the biological mechanisms of LTP and LTD have been performed in the hippocampus, particularly at the synapses where axons from cornu ammonis 3 (CA3) pyramidal cells contact the dendrites of pyramidal cells in CA1 via the Schaffer collateral pathway (figure 7.1*A*). In these experiments, commonly performed in hippocampal slices, Schaffer collaterals are stimulated and fEPSPs are recorded from the apical dendrites of CA1 pyramidal cells. A common protocol for inducing LTP in CA1 is the application of multiple trains of high-frequency stimulation (HFS), typically 100 Hz, while low-frequency stimulation (LFS), consisting of 1 Hz stimulation, is commonly used to elicit NMDA receptor-dependent LTD. The existing data suggest that many of the properties of Schaffer collateral-CA1 synapses are common to synapses throughout the neocortex, so in the following paragraphs we will describe the mechanisms for the best characterized and most common forms of LTP and LTD at synapses in area CA1.

The key requirement for Hebb-like synaptic modifications is the *coincident* activation of pre- and postsynaptic neurons; that is, cells must fire together to wire together. How, then, does one cell know when another cell is driving it to fire? Fast excitatory synaptic transmission at Schaffer collateral-CA1 synapses, along with most synapses in the central nervous system, is mediated primarily by glutamate-gated ion channels (i.e., glutamate receptors) embedded in the postsynaptic cell membrane. Glutamate receptors contribute to the fast excitatory postsynaptic response by allowing positively charged ions to flow into the postsynaptic compartment upon binding glutamate. Different classes of receptors are named for their most potent chemical agonists; in this chapter we will focus on AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole proprionate) receptors and NMDA (*N*-methyl-D-aspartate) receptors (figure 7.2*A*). In the mature brain, AMPA receptors primarily conduct Na^+ ions and mediate the initial and largest component of fEPSPs, while NMDA receptors have the unique property of allowing Ca^{2+} to flow into the postsynaptic cell in addition to Na^+. The contribution of NMDA receptors to basal synaptic transmission is highly variable because, under resting physiological conditions, their ion pores are blocked by Mg^{2+} ions which prevent current flow. However, during periods of heightened stimulation, the amount of positive charge flowing in through AMPA receptors depolarizes the postsynaptic spine sufficiently to remove the Mg^{2+} block and allow Ca^{2+} to enter the postsynaptic terminal through the NMDA receptor (middle panel, figure 7.2*A*) (Mayer, Westbrook, & Guthrie, 1984; Nowak, Bregestovski, Ascher, Herbet, & Prochiantz, 1984). The NMDA receptor thus satisfies the requirement of Hebb's coincidence detector, with its activation requiring *concurrent* presynaptic glutamate release and postsynaptic depolarization. The amount of Ca^{2+} passing through the NMDA receptor varies depending on the magnitude of pre- and postsynaptic coactivation and ultimately determines whether synaptic connections will be strengthened or weakened (Cummings, Mulkey, Nicoll, & Malenka, 1996; for review, see Lisman, 1985; Madison, Malenka, & Nicoll, 1991; Malenka & Nicoll, 1999). The most commonly studied forms of LTP and LTD are NMDA receptor dependent.

What are the physical and chemical changes that bring about the strengthening and weakening of synapses? The induction of LTP requires large influxes of Ca^{2+} in the postsynaptic terminal that trigger the activation of *protein kinases*, enzymes that add a phosphate group to specific proteins and modify their function, and whose enzymatic activity can long outlast the changes in Ca^{2+} concentration that triggered them (figure 7.2*A* right panel, figure 7.2*B*). The phosphorylation of synaptic and structural substrates, such as AMPA receptors and synaptic scaffolding proteins, plays a critical role in the induction and expression of LTP (see Browning, Huganir, & Greengard, 1985; Soderling & Derkach, 2000, for review). In addition to the phosphorylation of AMPA receptors already at synapses, new AMPA receptors are also inserted in the postsynaptic membrane (figure 7.2*A*, right panel) (Hayashi et al., 2000; Heynen, Quinlan, Bae, & Bear, 2000; Shi et al., 1999; see Malinow, Mainen, & Hayashi, 2000, for review), leading to larger amplitude fEPSPs with faster onsets, while on the presynaptic side more glutamate is released

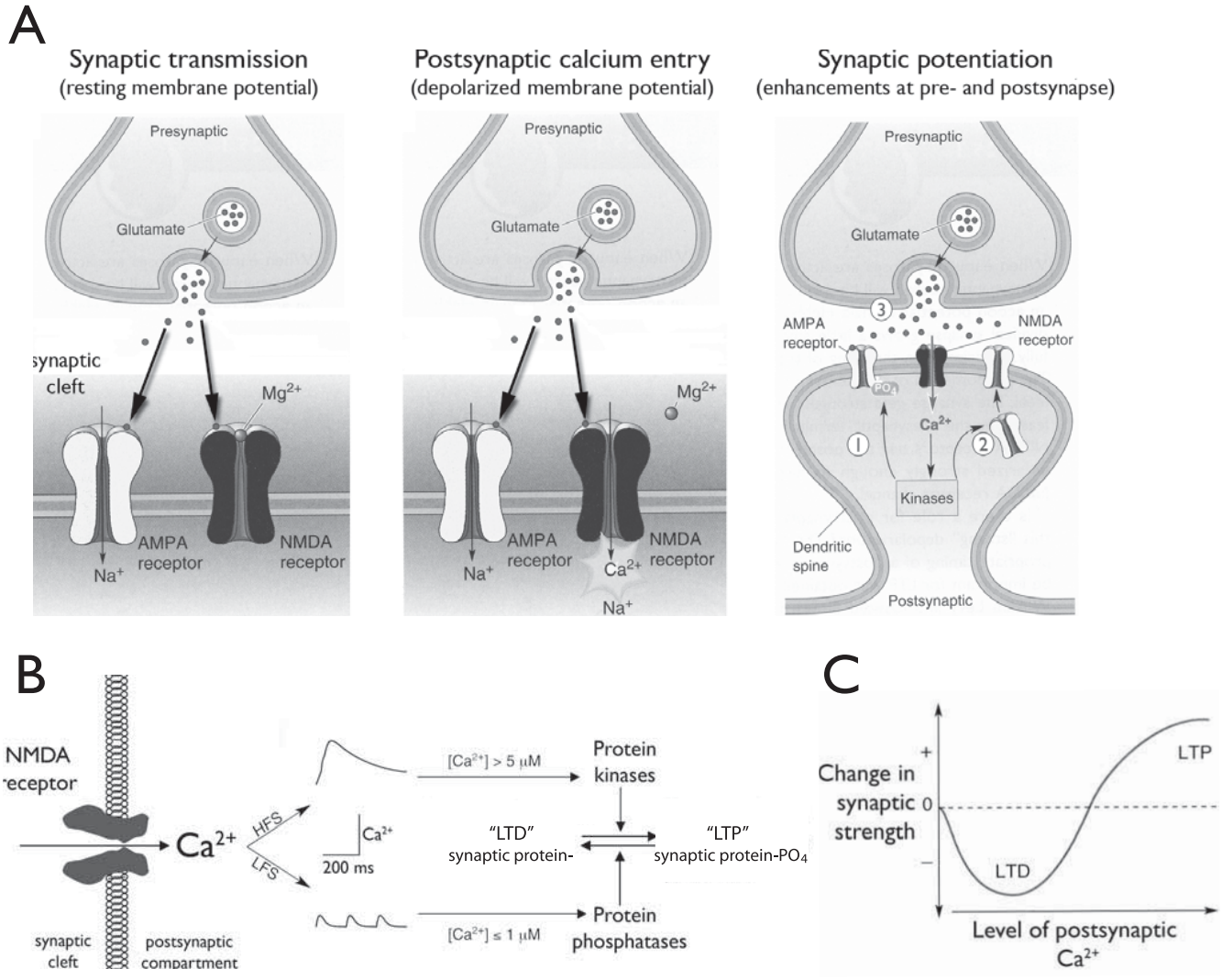


FIGURE 7.2 Postsynaptic calcium entry is the key for inducing LTP and LTD. (*A*) The NMDA receptor is activated by coincident pre- and postsynaptic activity. (*left*) During synaptic transmission, glutamate is released into the synaptic cleft and acts on AMPA and NMDA receptors, though NMDA receptors are blocked by Mg^{2+} ions at negative (resting) membrane potentials; (*middle*) if glutamate release coincides with sufficient postsynaptic depolarization, the Mg^{2+} block is removed and Ca^{2+} enters the postsynaptic neuron through NMDA receptors; (*right*) postsynaptic kinases initiate synaptic potentiation by (1) phosphorylating AMPA receptors already at the synapse, and (2) driving additional AMPA receptors to synapses. Glutamate release into the synaptic cleft (3) is also enhanced. (*B*) A large, brief increase in postsynaptic calcium, induced here by high-frequency stimulation (HFS), favors the activation of protein kinases and results in LTP, while small, sustained Ca^{2+} elevations during low-frequency stimulation (LFS) favor the activation of protein phosphatases, resulting in the dephosphorylation of synaptic proteins and LTD. (*C*) Long-term changes in synaptic strength can be explained as a function of the amount of calcium flowing into the postsynaptic neuron via NMDA receptors. (Modified with permission from M. Bear, B. Connors, & M. Paradiso, *Neuroscience: Exploring the brain*, 3rd edition, © 2007, Lippincott Williams & Wilkins.)

into the synaptic cleft (Bliss, Errington, & Lynch, 1990; Bliss, Errington, Lynch, & Williams, 1990; Dolphin, Errington, & Bliss, 1982). Structural changes, such as the growth of new spines and the enlargement or splitting of synapses in two, have also been observed following LTP induction (Abraham & Williams, 2003; Chen, Rex, Casale, Gall, & Lynch, 2007; Nagerl, Eberhorn, Cambridge, & Bonhoeffer, 2004; see Yuste & Bonhoeffer, 2001, for review). LTP lasting several hours or days (sometimes referred to as L-LTP) requires the synthesis of new proteins, and will gradually decay back to baseline if protein synthesis inhibitors are applied within the first couple of hours following conditioning stimulation (Frey, Krug, Reymann, & Matthies, 1988; Krug, Lossner, & Ott, 1984; Stanton & Sarvey, 1984; see Kelleher, Govindarajan, & Tonegawa, 2004, for review).

In the case of LTD, small increases in Ca^{2+} arising from weak synaptic stimulation favor the activation of *protein phosphatases* that dephosphorylate synaptic proteins including glutamate receptors (figure 7.2*B*) (Mulkey, Endo, Shenolikar, & Malenka, 1994; Mulkey, Herron, & Malenka,

1993). Contrary to LTP, LTD results in the removal and eventual degradation of AMPA receptors, NMDA receptors, and structural synaptic proteins (Ehlers, 2000; Heynen et al., 2000; Colledge et al., 2003), as well as the retraction of existing spines (Nagerl et al., 2004; Zhou, Homma, & Poo, 2004). Despite the fact that LTD involves the destruction of some preexisting proteins, long-lasting LTD (i.e., L-LTD), like L-LTP, also depends on the synthesis of new proteins—in fact, recent experiments have shown that proteins synthesized in response to L-LTP induction at one set of synapses can also be used to sustain L-LTD at nearby synapses on the same cell (Sajikumar & Frey, 2004). For a lengthier description of the mechanisms of LTP and LTD we recommend reviews by Bliss and Collingridge (1993), Malenka and Bear (2004), and Malinow and Malenka (2002).

Properties of LTP and LTD that are relevant for memory formation

Many of the physiological properties of LTP and LTD are homologous to the characteristics of behavioral memory expressed at the level of the whole animal, such as rapid induction (enabling fast learning) and longevity (allowing some memories to last a lifetime). While LTP and LTD are generally accepted as the leading cellular mechanisms for learning and memory, it should be noted that certain aspects of memory do not necessarily translate directly from changes at synapses, but more likely emerge at the level of the neural network in which the synaptic modifications are embedded (Hebb, 1949; Marr, 1971). The functional features exhibited by LTP and LTD are exactly the type that would be useful in enabling neural networks to rapidly acquire and store large amounts of information during behavior.

First, induction is rapid and long lasting. Changes in synaptic strength can be induced following very brief trains of high-frequency stimulation (Douglas & Goddard, 1975), and the resulting potentiation can last anywhere from several minutes to perhaps the entire lifetime of an animal (Abraham, Logan, Greenwood, & Dragunow, 2002; Barnes, 1979). These properties allow information learned from very brief episodes to be remembered for a lifetime, such as to not stick one's finger in a light socket.

Second, LTP provides a cellular mechanism for association of inputs to different synapses of a cell. This is indirectly apparent from the fact that the probability of inducing LTP increases with the number of stimulated afferents, a phenomenon referred to as *cooperativity* (McNaughton, Douglas, & Goddard, 1978). Weak stimulation will only affect a small proportion of synapses and is less likely to induce a postsynaptic change, whereas a strong stimulus will affect more synapses and increase the likelihood of inducing a long-lasting change in the postsynaptic response (figure 7.3, left). Transiently increasing the stimulation intensity during the delivery of a tetanus will lead to the recruitment of additional afferents and cause potentiation of synapses that would not have been coactivated during a weaker stimulation. In this sense, "cooperativity" is a form of associative synaptic potentiation (discussed in the next paragraph).

A more direct illustration of *associativity* involves the observation that when both weak and strong inputs are stimulated together, the weak input will show LTP, whereas if the weak input is stimulated alone, no LTP will be seen (figure 7.3, middle) (Barrionuevo & Brown, 1983; Levy & Steward, 1979). Associativity is relevant to learning and memory because it allows neurons to associate arbitrary patterns of activity from distinct neural pathways that may relay information regarding distinct but related events. This

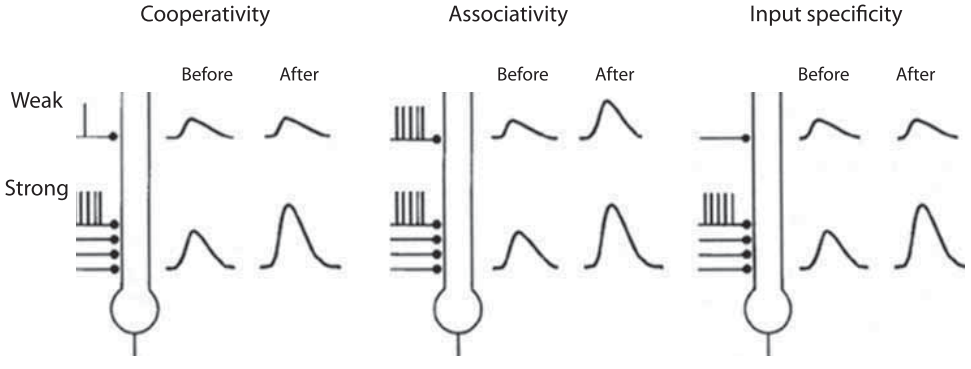


Figure 7.3 LTP exhibits physiological properties that make it a tenable cellular substrate for memory. Cooperativity (*left*) describes the property whereby a weak tetanus that activates relatively few afferents will not induce a change in the synaptic response, whereas coactivating many inputs with sufficiently strong stimulation will induce a change. Associativity (*middle*) describes the property whereby the concurrent simulation of weak and strong convergent pathways results in the long-term strengthening of the weak pathway. (*right*) LTP is input specific because only those synapses active at the time of the tetanus express potentiation; inactive inputs do not share in the potentiation. Cooperativity, associativity, and input specificity apply similarly to LTD. (Modified with permission from R. Nicoll, J. Kauer, & R. Malenka, The current excitement in long-term potentiation, *Neuron*, *1*, 97–103, © 1988, Cell Press.)

phenomenon is epitomized by recent experiments showing that two populations of synapses given either weak or strong stimulation both express a biochemical "plasticity tag" that allows them to capture "plasticity-related proteins" synthesized specifically in response to the strong stimulation (Frey & Morris, 1997). This process, known as *synaptic tagging*, allows weakly stimulated synapses to sequester and utilize the same proteins as strongly stimulated synapses and express long-lasting LTP as though they had received strong stimulation.

The property of NMDA receptor-dependent LTP and LTD perhaps most relevant to memory formation is input specificity, where only the synaptic pathways that receive conditioning stimulation show a change in synaptic strength; other synapses on the same postsynaptic cell that do not receive stimulation do not express plasticity (figure 7.3, right) (Andersen et al., 1977; Bliss & Lømo, 1973; Dudek & Bear, 1992). This property is relevant to learning and memory for two reasons: (1) because it ensures that only the particular synapses that were activated by an experience will store information relevant to that experience (of what use is LTP as a memory mechanism if one event results in the nonspecific potentiation of all the synapses in your brain?), and (2) because the information storage capacity of a neuron is much greater when information is encoded at individual synapses rather than as a change in a property of the entire cell (such as whole-cell excitability).

Studies of LTP and LTD in behaving animals

The functional properties of LTP and LTD are suggestive of an involvement in memory processes, but how can this be tested in behaving animals? Over the years, researchers have proposed different strategies for testing a necessary and sufficient role for synaptic modification in behavioral learning (Martin, Grimwood, & Morris, 2000; Neves, Cooke, & Bliss, 2008; Stevens, 1998). *Anterograde alteration* and *retrograde alteration* approaches, historically the most common, seek to describe correlations between synaptic plasticity and behavioral memory by manipulating various mechanisms of LTP or LTD before or after learning and characterizing the ensuing changes in learning or memory maintenance; the goal of the *detection* strategy is to determine whether LTP- or LTD-like changes are induced naturally as a consequence of learning; and *mimicry*, the most technically challenging and hitherto unrealized approach, would seek to engineer a memory of an experience that never actually occurred by tweaking just the relevant synaptic weights (Neves et al., 2008). Successfully installing an "artificial" memory in this manner would provide the ultimate proof that changes in synaptic weight are sufficient for memory storage, but because this has not yet happened we shall focus instead on the first three strategies.

Anterograde Alteration Studies in recent years have demonstrated unequivocally that many of the mechanisms necessary for the induction and maintenance of LTP are also necessary for the acquisition and maintenance of memory (though there are arguments rebutting the notion that LTP is involved in memory; see Shors & Matzel, 1997, for review). The basic assumption of the anterograde alteration strategy holds that, if learning new information requires the induction of LTP, then blocking LTP prior to learning should prevent memory formation. One of the first pharmacological studies linking LTP induction to spatial learning involved infusing the selective NMDA receptor antagonist APV into the lateral ventricles of rats prior to learning the *Morris water maze* task, in which animals learn to swim through opaque water to find a submerged escape platform using visual cues outside the water tank as guides (figure 7.4*A*) (Morris, Anderson, Lynch, & Baudry, 1986). At first the animals swim randomly through the maze until they happen to find the platform, but they eventually learn to swim directly to the correct location. In the study by Morris and colleagues, rats treated with a dose of APV sufficient to block LTP induction in vivo were impaired at learning the water maze task and showed no preference for the target quadrant after the platform was removed, whereas animals treated with the inactive isoform of the drug performed the same as saline-treated controls (figure 7.4*B*). Furthermore, subsequent work showed that the extent of the spatial learning deficit correlated with the degree of LTP impairment as the dose of APV was increased (Butcher, Davis, & Morris, 1990). Some criticisms raised against these pharmacological studies, however, were that APV infusions caused sensory and motor deficits unrelated to learning (Abraham & Kairiss, 1988; Caramanos & Shapiro, 1994; Shors & Matzel, 1997) and that animals given pretraining in different mazes were able to learn new platform locations even when NMDA receptors were blocked (Bannerman, Good, Butcher, Ramsay, & Morris, 1995; Saucier & Cain, 1995). These concerns were addressed by a follow-up study demonstrating that direct hippocampal infusions of APV caused a delay-dependent impairment in rapidly acquired spatial learning that could not be explained by interference from drug side effects (Steele & Morris, 1999).

An essential role for NMDA receptor-dependent synaptic plasticity in spatial memory was further demonstrated through the use of mice with a regionally restricted, postnatal genetic deletion of the obligatory NMDA receptor subunit NR1 (Tsien, Huerta, & Tonegawa, 1996). This study represented an important technological advance over previous studies using knockout mice because the spatial and temporal restrictions of the genetic deletion ruled out compensatory side effects during development as a possible explanation for any learning deficits. It was found that deleting the NR1 gene exclusively in CA1 pyramidal cells led to the specific

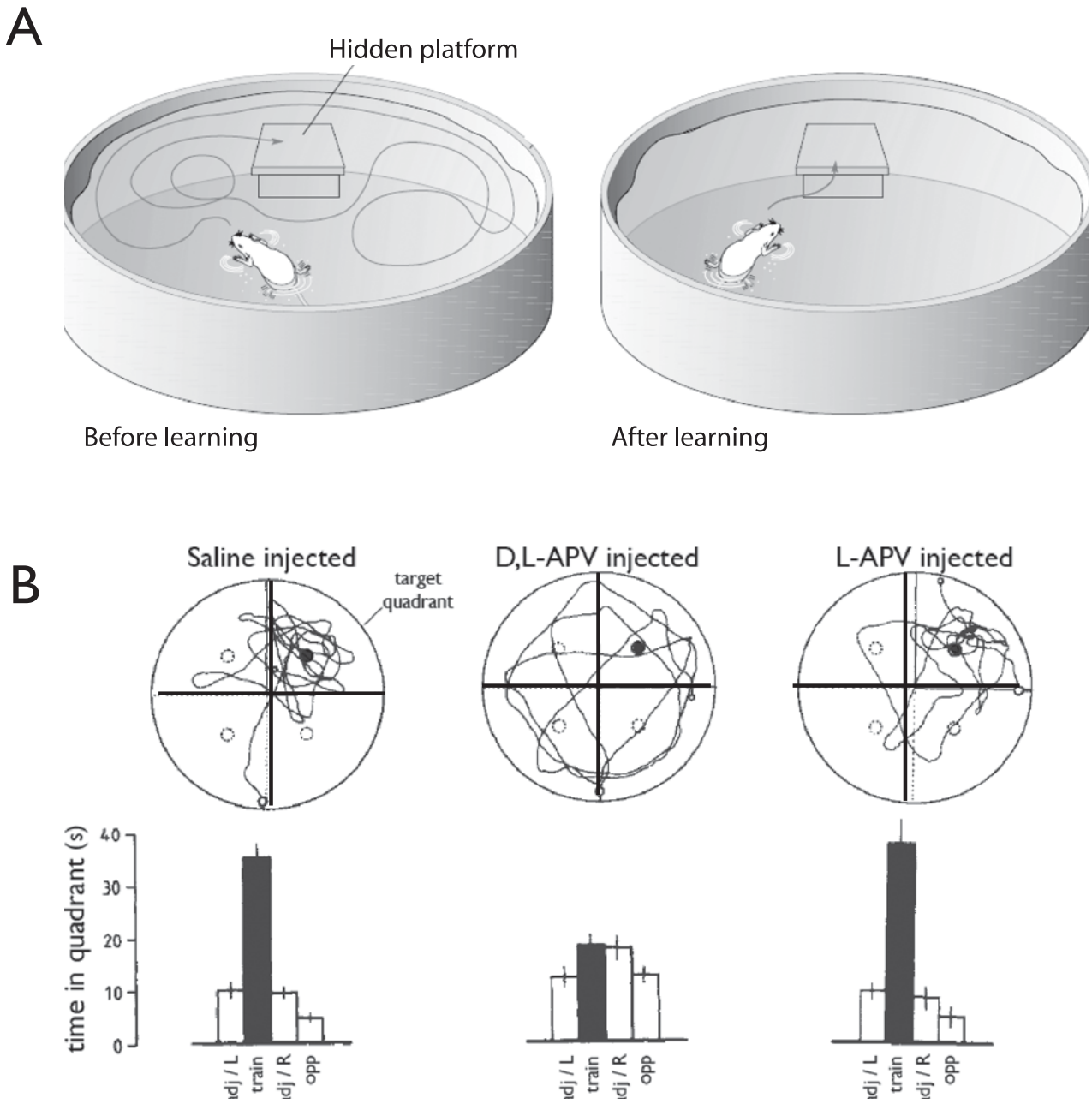


FIGURE 7.4 The Morris water maze is one of the most common behavioral tests of spatial learning and memory. (*A*) In the task, an animal is placed in a pool filled with opaque water and must locate a submerged escape platform. At the start of training the animal's swim path is typically long and circuitous. After several training trials the rat learns the platform location and will swim straight to it during a test trial. (Modified with permission from M. Bear, B. Connors, & M. Paradiso, *Neuroscience: Exploring the brain*, 3rd edition, © 2007, Lippincott Williams & Wilkins.) (*B*) Shown at top are the swim trajectories of animals tested in the Morris water maze after having been given different drug treatments prior to training. Animals injected with saline (left) spent the greatest amount of time in the target quadrant (which earlier had the escape platform), whereas animals treated with the NMDA receptor antagonist D,L-APV (middle) showed no preference for the target quadrant. Animals injected with the inactive L-isomer of APV (right) showed normal spatial memory similar to the saline-injected group. (Modified with permission from Morris, Anderson, Lynch, & Baudry, Selective impairment of learning and blockade of long-term potentiation by an *N*-methyl-D-aspartate receptor antagonist, AP5, *Nature*, *319*, 774–776, © 1986, *Nature* [Nature Publishing Group].)

ablation of LTP in CA1, while LTP in the dentate gyrus was unaffected. Furthermore, CA1-knockout mice showed no place preference for the target quadrant in the test phase of the Morris water maze task. Subsequent studies have refined the temporal resolution of genetic manipulations even more through the use of genetically modified transcription factors that can be reversibly activated or inactivated over the course of a few days by interacting with innocuous drugs administered in the animals' food (Mansuy et al., 1998; Mayford & Kandel, 1999).

NMDA receptor activation is also necessary for the acquisition of associative Pavlovian learning tasks, such as contextual fear conditioning, where animals learn to associate a particular context with a foot shock, and cued fear conditioning, in which animals learn that the presentation of an auditory tone predicts an imminent foot shock (see Maren, 2001, for review). Similarly, inhibitory avoidance training, a single-trial learning paradigm in which animals learn to avoid the shock-associated compartment of a conditioning apparatus, depends on NMDA receptor activation (Izquierdo et al., 1992; Jerusalinsky et al., 1992). A totally different approach to assess the role of LTP in learning and memory is to drive LTP in a hippocampal pathway to saturation prior to learning (McNaughton, Barnes, Rao, Baldwin, & Rasmussen, 1986; Moser, Krobert, Moser, & Morris, 1998). The hypothesis is that, if information is stored as increases in synaptic weights, then saturating LTP at hippocampal synapses should impair subsequent learning. To achieve this purpose, the investigators implanted multiple stimulating electrodes into the perforant path and applied several trains of HFS to drive LTP in the dentate gyrus to asymptote. Animals in which LTP was truly saturated were impaired at learning the Morris water maze, whereas other animals in which as little as 10% residual LTP could be induced were able to learn normally (Moser et al.). This observation suggests that even a very small capacity for LTP is sufficient for the hippocampus to store spatial information.

Retrograde Alteration Another strategy for assessing the involvement of LTP in memory is to compromise the expression or maintenance of LTP after learning. This strategy is predicated on the assumption that, if information is stored as distributed modifications in synaptic connections, then memory should be susceptible to disruption by manipulations that alter or erase the pattern of synaptic weights (figure 7.5*A*). A successful example comes from a follow-up study to the 1998 work by Moser and colleagues, where the approach was to induce LTP in the dentate gyrus *after* animals were trained in the Morris water maze to compromise the learning-induced pattern of synaptic weights (Brun, Ytterbo, Morris, Moser, & Moser, 2001). Indeed, posttraining LTP induction impaired the rats' memory of the platform location. A separate group of animals given an NMDA receptor antagonist after learning but before LTP induction expressed normal spatial memory, suggesting that the memory impairments were generated specifically as a consequence of the artificial manipulation of synaptic weights.

In addition to the electrophysiological approach, researchers have used pharmacological agents to block specific kinases necessary for LTP induction and maintenance to disrupt memory. For example, infusing inhibitors of calcium/calmodulin-dependent kinase II (CaMKII) or protein kinase C (PKC), enzymes critical for the establishment of LTP, into the hippocampus of rats soon after inhibitory avoidance training was found to cause full retrograde amnesia in animals tested 24 hours later (Paratcha et al., 2000; Wolfman et al., 1994). It is not known whether the amnesia was caused by a reversal of LTP in these studies, but more recent work has reported a parallel erasure of LTP and hippocampus-dependent spatial memory. Pastalkova and colleagues infused an inhibitor of the zeta isoform of protein kinase M (PKMζ), an enzyme whose constitutive activity is necessary for the maintenance of long-term hippocampal LTP in vivo (figure 7.5*B*) (Pastalkova et al., 2006). The fact that the drug simply reversed preexisting LTP without apparently affecting synaptic transmission or subsequent LTP induction in any other way made it ideal for testing the importance of LTP in long-term memory maintenance. In the study, rats learned to associate a particular area on a rotating platform with a mild foot shock, and acquired a robust avoidance response after just a few training trials. Infusing the PKMζ inhibitor into the hippocampus 24 hours after training eradicated the avoidance response (figure 7.5*C*), and, shockingly, had the same effect when rats were treated *1 month* after learning. The results challenged the view that protein kinases play only a time-limited role in LTP and memory formation, and that lifelong memories are eventually stored independently of the hippocampus.

Detection Before it can be accepted that learning and LTP modify synapses in the same way, it must be demonstrated that learning causes detectable changes in synaptic strength that (1) resemble and (2) occlude LTP. There are a growing number of studies which report learning-related enhancements in synaptic responses in areas of the brain appropriate to the type of information learned. We will briefly elaborate on a few of the studies demonstrating long-lasting synaptic modifications occurring as a result of learning.

Some of the clearest electrophysiological evidence linking neocortical LTP and learning has come from studies using skill learning in rats, where the animals were trained to reach through a small window and retrieve a food pellet with their forepaw. When recordings were later conducted on brain slices prepared from trained animals, it was found that fEPSPs

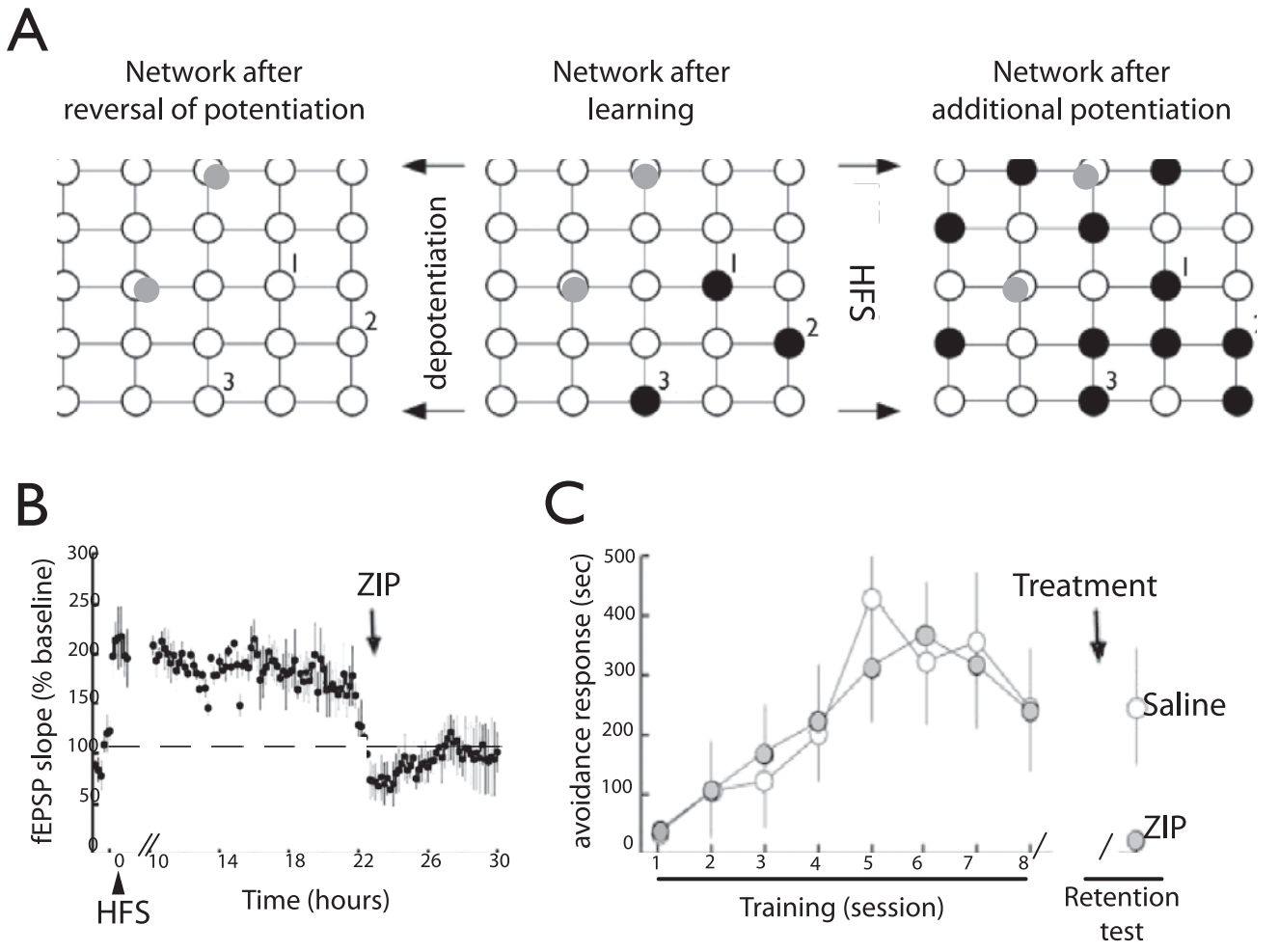


FIGURE 7.5 Disrupting the pattern of synaptic weights in a network results in a loss of the information stored across the synapses. (*A*) A hypothetical distribution of synaptic enhancements induced in a network by learning; lines are neuronal processes which intersect at synapses represented as circles; black circles are synapses potentiated by recent learning, gray circles are synapses already potentiated from an unrelated event; white circles are unpotentiated synapses. (*top right*) Randomly potentiating irrelevant synapses with high-frequency stimulation (HFS) after learning scrambles the pattern of learning-induced synaptic weights and disrupts memory storage; this is the experimental strategy used by Brun, Ytterbo, Morris, Moser, and Moser (2001). (*top left*) The reversal of learning-related synaptic enhancements should erase the information stored across the connections and cause retrograde amnesia. (Modified with permission from Brun et al., Retrograde amnesia for spatial memory induced by NMDA receptor-mediated long-term potentiation, *Journal of Neuroscience*, *21*(1), 356–362, © 2001, Society for Neuroscience.) (*B*) Robust LTP induced in the dentate gyrus in vivo can be rapidly reversed 22 hours later by intrahippocampal infusion of the selective PKMζ antagonist "zeta inhibitory peptide," or ZIP. (*C*) In parallel with the reversal of LTP, intrahippocampal infusions of ZIP caused abrupt and complete amnesia in a place-avoidance task in rats tested either 24 hours or 1 month after training. The avoidance memory of saline-infused animals remained intact. (Modified with permission from Pastalkova et al., Storage of spatial information by the maintenance mechanism of LTP, *Science*, *313*, 1141–1144, © 2006, American Association for the Advancement of Science.)

in the primary motor cortex (M1) corresponding to the preferred reaching paw were substantially larger than fEPSPs from the hemisphere for the untrained paw (i.e., the "untrained" hemisphere) (figure 7.6*A*) (Rioult-Pedotti, Friedman, Hess, & Donoghue, 1998). Follow-up studies investigated the impact of skill learning on subsequent LTP and LTD and revealed a marked reduction in the amount of LTP and an enhancement in the magnitude of LTD in the "trained" portion of cortex (figure 7.6*B*) (Monfils & Teskey, 2004; Rioult-Pedotti, Friedman, & Donoghue, 2000), suggesting that learning had elevated the synapses in motor cortex closer to their ceiling for LTP expression and, concurrently, left more space for synaptic depression. The partial reduction, or "occlusion," of LTP by learning suggests that skill learning and LTP engage a common neural mechanism. More recent work has shown that the synaptic modification range shifts upward to accommodate the synaptic enhancements a few weeks after learning, thereby restoring the capacity of the connections to express their previous levels of LTP and LTD (Rioult-Pedotti, Donoghue, & Dunaevsky, 2007).

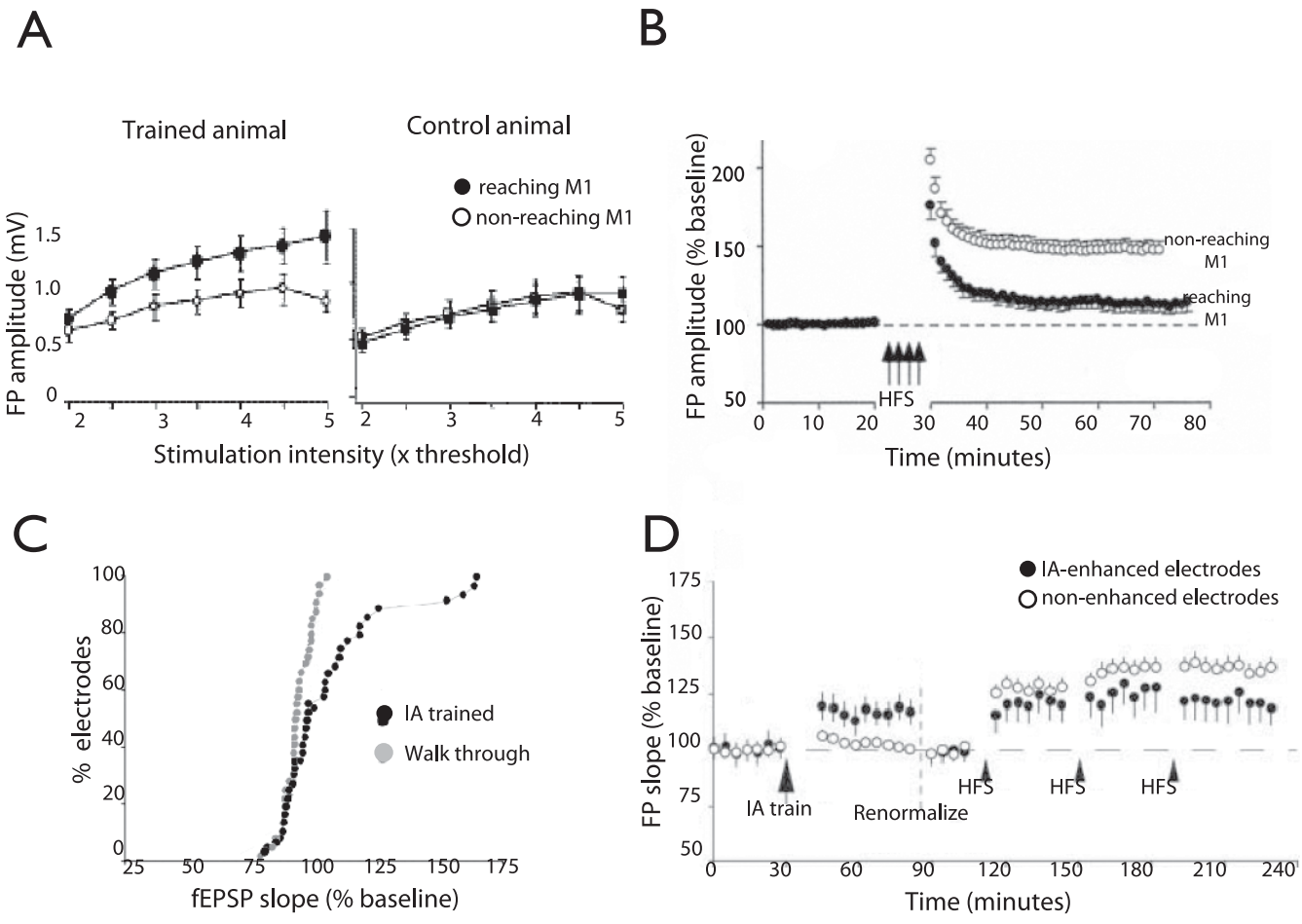


FIGURE 7.6 Learning induces synaptic enhancements that occlude LTP in brain areas relevant to the type of information learned. (*A*) Learning a new motor skill enhanced fEPSP amplitude specifically in the forelimb region of primary motor cortex (M1) corresponding to the preferred reaching paw in trained rats; no enhancements were observed in the same area of M1 in untrained control animals. (Modified with permission from Rioult-Pedotti, Friedman, Hess, & Donoghue, Strengthening of horizontal cortical connections following skill learning, *Nature Neuroscience*, *1*(3), 230–234, © 1998, *Nature* [Nature Publishing Group].) (*B*) The fEPSP enhancements associated with skill learning resulted in the partial occlusion of LTP in M1 in the "reaching" hemisphere compared to the "nonreaching" hemisphere in trained rats. (Modified with permission from Rioult-Pedotti, Friedman, & Donoghue, Learning-induced LTP in neocortex, *Science*, *290*, 533–536, © 2000, American Association for the Advancement of Science.) (*C*) In vivo recording experiments in rats revealed that single-trial inhibitory avoidance (IA) training led to fEPSP enhancements in a subpopulation of recording electrodes in the hippocampus of trained animals relative to controls (who walked through the training apparatus without receiving a foot shock, i.e., the "Walk through" group). Data were collected 2 hours after conditioning. (*D*) Electrodes showing fEPSP enhancements upon IA training reached LTP saturation more rapidly and showed less LTP in response to repeated trains of HFS, demonstrating that this form of learning mimicked and occluded hippocampal LTP in vivo. (Modified with permission from Whitlock, Heynen, Shuler, & Bear, Learning induces long-term potentiation in the hippocampus, *Science*, *313*, 1093–1097, © 2006, American Association for the Advancement of Science.)

Studies in the amygdala have also yielded strong evidence linking LTP and memory formation. Because of its well-characterized anatomical connectivity, the amygdala has allowed neuroscientists the opportunity to directly investigate associative synaptic plasticity between distinct inputs following associative learning. One of the most common experimental approaches has been to use Pavlovian fear-conditioning paradigms in which the aversive, fear-evoking stimulus of a foot shock (the US) is paired with a novel environmental cue, such as a tone (the CS). Research in the 1990s showed that repeatedly pairing a tone with a foot shock resulted in the strengthening of auditory thalamic inputs to the amygdala and increases in the amplitude of auditory-evoked responses when animals were replayed the tone after conditioning (McKernan & Shinnick-Gallagher, 1997; Rogan, Staubli, & LeDoux, 1997). Thus the initially weak tone representation became potentiated through its association with the foot shock. Similar to LTP, it was found that this form of learning resulted in the delivery of AMPA receptors to amygdalar synapses and that blocking

the synaptic delivery of AMPA receptors prevented the formation of the fear memory (Rumpel, LeDoux, Zador, & Malinow, 2005).

In addition to the amygdala, several studies have demonstrated learning-specific, LTP-like changes in the synaptic expression and phosphorylation of AMPA receptors in the hippocampus following tasks such as contextual fear conditioning, where animals learn to fear a context as opposed to a discrete tone, and inhibitory avoidance training (Shukla, Kim, Blundell, & Powell, 2007; Matsuo, Reijmers, & Mayford, 2008; Cammarota, Bernabeu, Levi De Stein, Izquierdo, & Medina, 1998; Bevilaqua, Medina, Izquierdo, & Cammarota, 2005; Whitlock, Heynen, Shuler, & Bear, 2006). Many of the downstream biochemical cascades initiated by inhibitory avoidance training are the same as those seen following LTP induction in the hippocampus (for review, see Izquierdo et al., 2006). Electrophysiological experiments have further confirmed the occurrence of LTP-like modification of hippocampal synapses following various learning tasks. In one such study, Sacchetti and colleagues showed that hippocampal slices obtained from rats after contextual fear conditioning showed fEPSP enhancements that partially occluded subsequent LTP, suggesting that contextual learning and LTP shared a common expression mechanism (Sacchetti et al., 2001, 2002). LTP-like enhancements in synaptic transmission have also been recorded in area CA1 following trace eyeblink conditioning, a form of hippocampal-dependent associative learning. Enhanced fEPSP responses were reported following this task in hippocampal slices prepared from trained rabbits (Power, Thompson, Moyer, & Disterhoft, 1997), as well as in the intact hippocampus of freely behaving mice (Gruart, Munoz, & Delgado-Garcia, 2006). Some of the most conclusive evidence demonstrating learning-induced LTP in the hippocampus comes from recent in vivo recording experiments in rats that were given inhibitory avoidance training (Whitlock et al., 2006). Multielectrode arrays were chronically implanted to record fEPSPs at several sites in the hippocampus of awake, behaving animals before and after inhibitory avoidance training. The training caused abrupt and long-lasting (>3 hr) enhancements of evoked fEPSPs in a subpopulation of the recording electrodes in trained animals relative to controls (figure 7.6*C*). Additional experiments demonstrated that electrodes showing training-related fEPSP enhancements expressed less subsequent LTP in response to HFS than neighboring electrodes that were not enhanced by training—that is, the learning-related enhancements partially occluded subsequent LTP (figure 7.6*D*). This demonstration of learning-induced fEPSP enhancements that occlude LTP in vivo provided long-awaited evidence that the strengthening of hippocampal synapses is a natural physiological occurrence following some forms of associative learning.

How does LTP influence hippocampal receptive fields?

If changes in synaptic transmission ultimately result in modified behavior, then they must change the way in which the brain structures that mediate those behaviors communicate with one another. A mechanistic understanding of how LTP contributes to behavior therefore requires a description of how changes in synaptic strength affect representations in neural networks. A well-studied experimental tool for understanding neural representations has been hippocampal place cells, first characterized in area CA1, which discharge only when an animal occupies a particular spatial location, the "place field" (O'Keefe & Dostrovsky, 1971). Neighboring place cells express distinct but overlapping place fields such that the entire surface of a recording environment is completely represented by a group of cells (O'Keefe, 1976). The spatial representations of place cells are extremely specific, with the cells firing at entirely unrelated locations from one recording environment to the next (O'Keefe & Conway, 1978), and can be incredibly stable, maintaining the same firing field locations for as long as the cells are identifiable (Thompson & Best, 1990). More recent advances in recording technology have enabled researchers to simultaneously record the activity of large ensembles of place cells (>100) as new map representations emerged during exploration of a novel environment (Wilson & McNaughton, 1993). Because the concerted activity of large groups of cells in the hippocampus will completely cover any environment encountered by an animal, it has been hypothesized that the hippocampus provides the neural substrate for an integrative "cognitive map," which provides "an objective spatial framework within which the items and events of an organism's experience are located and interrelated" (O'Keefe & Nadel, 1978). The remarkable spatial specificity and stability of place cells make them ideal candidates for contributing to a spatial memory system. Long-term potentiation is thought to fit into this framework by providing a synapse-specific mechanism enabling the long-lasting storage of spatial representations for a potentially very large number of environments. In this section we review studies that have begun to establish a link between LTP and place cell representations.

One of the earliest studies demonstrating a mechanistic relationship between LTP and place fields used mice carrying a CA1-specific deletion of the gene encoding the NMDA receptor subunit NR1 (McHugh, Blum, Tsien, Tonegawa, & Wilson, 1996). In parallel with the previously mentioned impairments in LTP and spatial learning in these mice, the authors found that the firing fields of CA1 place cells exhibited somewhat reduced spatial specificity, although the place fields did not disappear entirely (i.e., they were broader and had less well-defined boundaries than control animals; figure 7.7*A*). Place cells with overlapping place fields also showed reduced covariance of firing, implying a reduced capacity

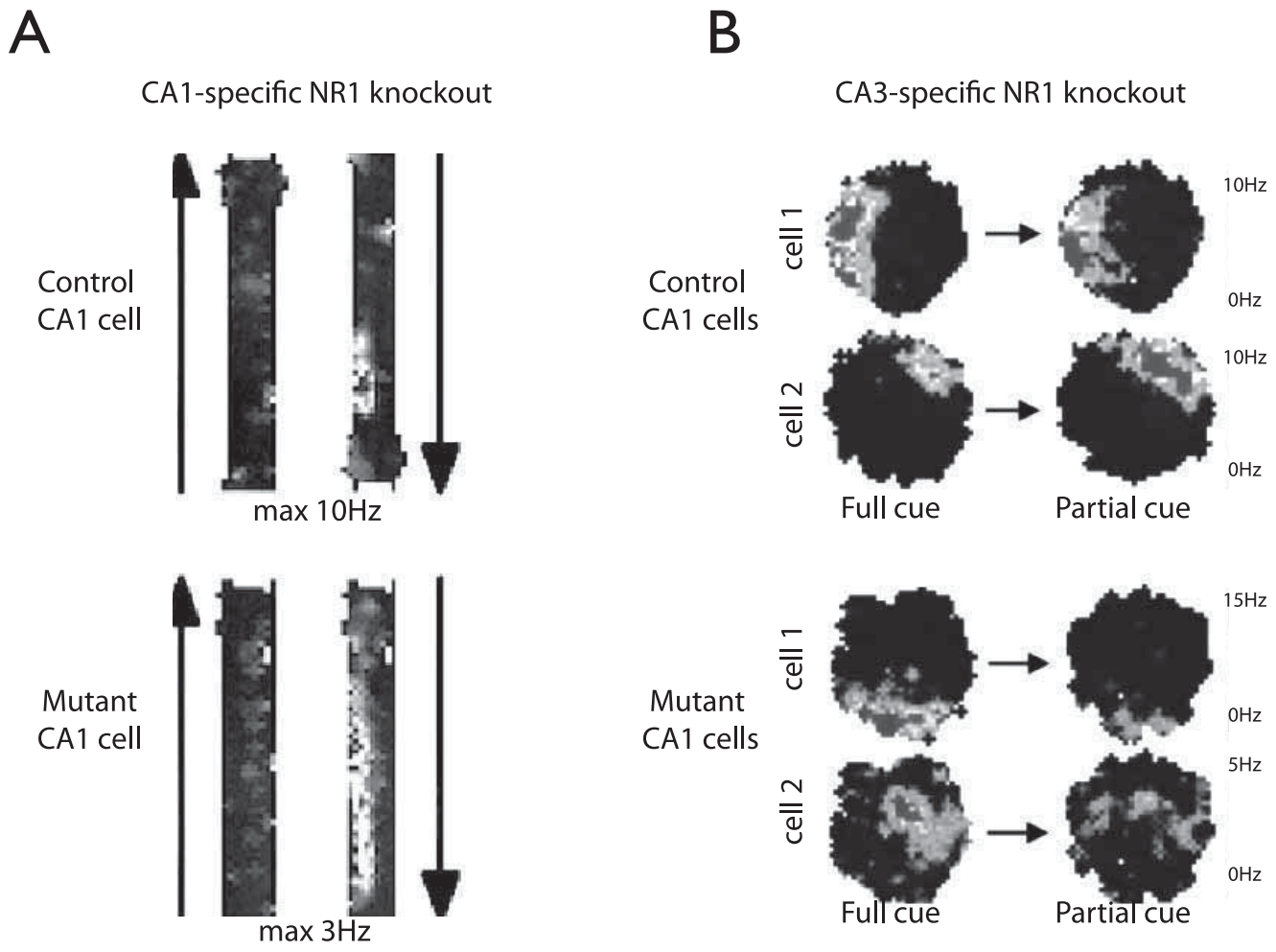


FIGURE 7.7 Genetic deletion of the obligatory NR1 subunit of the NMDA receptor alters place cell properties without preventing the expression of place fields per se. (*A*) Examples of direction-specific CA1 place cell activity from CA1-specific NR1 knockout mice and control mice running on a one-dimensional linear track; the panels show the firing rates of cells as a function of the location of the animals on the track. In this example, the cells were virtually silent when the animals traversed the track in the upward direction, but fired in a spatially restricted manner as the animals ran back down. Place fields in CA1 knockout mice were stable but significantly larger than in controls. (Modified with permission from McHugh, Blum, Tsien, Tonegawa, & Wilson, Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice, *Cell*, *87*, 1339–1349, © 1996, Cell Press.) (*B*) The firing fields of CA1 place cells in CA3-specific NR1 knockout mice differed from controls only during specific environmental manipulations. There were no differences in place field properties when four out of four distal cues were present in the recording arena ("full cue" condition); however, when mice were returned to the arena with only one of four cues present ("partial cue"), CA3 knockout mice expressed significantly smaller CA1 place fields and had lower firing rates than controls. These experiments suggested a functional role for CA3 in pattern completion. (Modified with permission from Nakazawa et al., Requirement for hippocampal CA3 NMDA receptors in associative memory recall, *Science*, *297*, 211–218, © 2002, American Association for the Advancement of Science.)

for coordinating ensemble codes for spatial location across cells. Considerable sparing of place-specific firing was also seen in mice with a specific deletion of the NR1 subunit in the CA3 subfield (Nakazawa et al., 2002). In these animals, the sharpness of firing fields in CA1 was not reduced at all. Impaired spatial firing appeared only under conditions where a substantial fraction of the landmarks in the environment were removed (figure 7.7*B*). Further insight was obtained in a study by Kentros and colleagues, where the authors used a pharmacological approach to compare the role of NMDA receptors in induction and maintenance of spatial representations (Kentros et al., 1998). They found that injecting rats with a selective NMDA-receptor antagonist prevented the maintenance of new place fields acquired in novel environments when animals were reexposed to the environments a day later. The drug treatment did not affect preexisting place fields in a familiar environment, and new place fields were expressed instantaneously as animals explored a novel environment. These observations, together with the spared spatial firing observed in the NR1 knockout

mice, suggest that the mechanism for the expression of place fields per se is NMDA-receptor independent. NMDA receptors may instead be necessary for maintaining place fields in fixed locations across different experiences in the environment.

NMDA receptor activation is also necessary for experience-dependent changes in place cell discharge properties, as revealed by experiments in which rats repeatedly traversed the length of a linear track. This behavior was associated with the asymmetric, backward expansion of CA1 place fields relative to the rat's direction of motion, which was hypothesized to aid in predicting elements of upcoming spatial sequences before they actually occurred (Mehta, Barnes, & McNaughton, 1997; Mehta, Quirk, & Wilson, 2000). This form of behaviorally driven receptive field plasticity was hypothesized to arise from LTP-like synaptic enhancements between cells in CA3 and CA1 and, indeed, the effect was blocked in animals injected with NMDA receptor antagonists (Ekstrom, Meltzer, McNaughton, & Barnes, 2001). In addition to studies demonstrating a permissive role for NMDA-receptor activation in place cell plasticity, evidence supporting an instructive role for LTP in driving changes in place representations comes from a study by Dragoi and colleagues. It was found that inducing LTP in the hippocampus caused remapping of place cell firing fields in familiar environments, including the creation of new fields, the disappearance of others, and changes in the directional preferences of others (Dragoi, Harris, & Buzsaki, 2003). Additional work revealed that contextual fear conditioning, which itself induces LTP-like enhancements of hippocampal fEPSPs (Sacchetti et al., 2001, 2002), also results in the partial remapping of place fields in CA1 (Moita, Rosis, Zhou, LeDoux, & Blair, 2004), suggesting that synaptic plasticity and place field plasticity are merely different aspects of a common mechanism engaged by the hippocampus during associative contextual learning.

Synaptic plasticity and attractor dynamics in neural networks

Place cells are thought to be part of a hippocampal system for storage of episodic memories with a spatial component (S. Leutgeb et al., 2005; S. Leutgeb, J. K. Leutgeb, Moser, & Moser, 2005). Several studies over the years have revealed that place cells encode more than just space, including odors, textures, temporal sequences, and prior events (Hampson, Heyser, & Deadwyler, 1993; Moita et al., 2004; Wood, Dudchenko, & Eichenbaum, 1999; Wood, Dudchenko, Robitsek, & Eichenbaum, 2000; Young, Fox, & Eichenbaum, 1994), and the place cell network is known to support a number of discrete and graded representations in the same environment (Bostock, Muller, & Kubie, 1991; J. Leutgeb et al., 2005; Markus et al., 1995; Muller & Kubie, 1987). This multiplicity of the hippocampal spatial map contrasts strongly with the universal nature of representations one synapse upstream, in the superficial layers of the medial entorhinal cortex, which interfaces most of the external sensory information from the cortex to the hippocampus and back (Fyhn, Molden, Witter, Moser, & Moser, 2004). The key cell type among the entorhinal inputs to the hippocampus is the grid cell, which fires at sharply defined locations like place cells in the hippocampus but differs from such cells in that each cell has multiple firing locations and that the firing locations of each cell form a tessellating triangular pattern across the entire environment available to the animal (Hafting, Fyhn, Molden, Moser, & Moser, 2005). Different grid cells have nonoverlapping firing fields; that is, the grids are offset relative to each other (Hafting et al., 2005), but the grids of different colocalized cells keep a constant spatial relationship between different environments, implying that a single spatial map may be used in all behavioral contexts, very much unlike the recruitment of discrete and apparently nonoverlapping representations in the hippocampus (Fyhn, Hafting, Treves, Moser, & Moser, 2007). These observations, taken as a whole, suggest that self-location is maintained and perhaps generated in entorhinal cortex (McNaughton, Battaglia, Jensen, Moser, & Moser, 2006), whereas the role of the hippocampus is to differentiate between places and experiences associated with places, and to associate each of them to the particular features of each environment. Such a regional differentiation would be consistent with a critical role for the hippocampus in memory for individual episodes.

But how are associative memories encoded and retrieved in the place cell system? It is commonly believed that memories are encoded at the level of neural ensembles and that the ensembles are implemented in neural attractor networks (Amit, Gutfreund, & Sompolinsky, 1985, 1987; Hopfield, 1982). An attractor network has one or several preferred positions or volumes in the space of network states, such that when the system is started from any location outside the preferred positions, it will evolve until it reaches one of the attractor basins (figure 7.8*A*). It will then stay there until the system receives new input. These properties allow stored memories (the preferred positions of the system) to be recalled from degraded versions of the original input (positions that are slightly different from the preferred positions). Storing different places and episodes as discrete states in such a network keeps memories separate and avoids memory interference. Attractor networks could, in principle, be hardwired, but for the hippocampus, as well as any other memory-storing system, this is quite unlikely considering that thousands of new memories are formed in the system each day. It is more likely that representations evolve over time, with new states being formed each time a new event is experienced. The formation of hippocampal attractor states is thought to

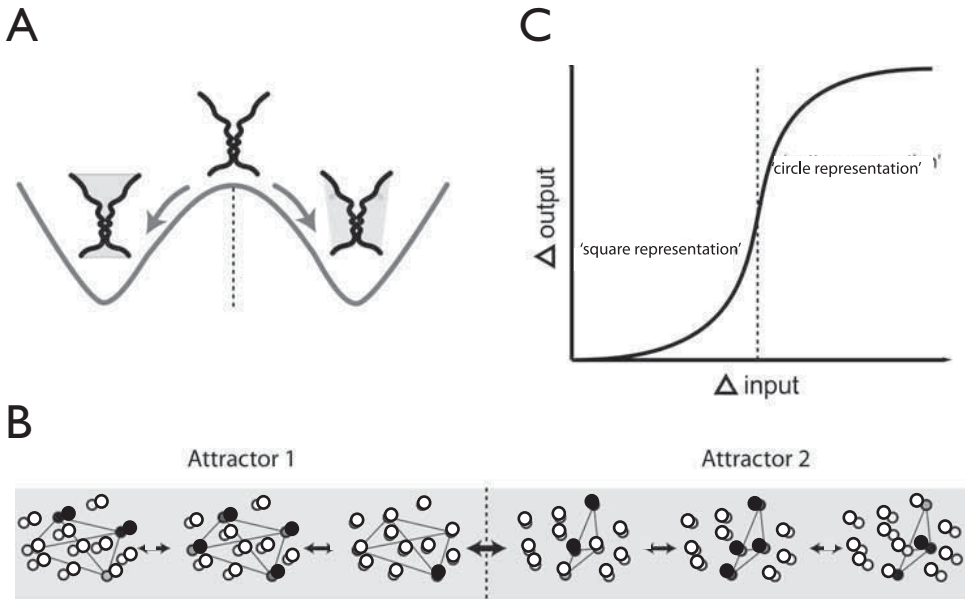


FIGURE 7.8 Patterns of activity in cell assemblies with attractor dynamics. (*A*) Ambiguous patterns tend to converge to a familiar matching pattern (i.e., pattern completion) and, simultaneously, diverge away from interfering patterns (i.e., pattern separation). This nonlinear process can be illustrated with an illusion using an ambiguous visual object. The perceived image in *A* will tend to fluctuate between two familiar images (a chalice on the left, or two kissing faces on the right), instead of stabilizing on the ambiguous white object in the middle. (*B*) The presence of attractor states in a neural network favors the emergence of familiar patterns even when the initial input is severely degraded. Activating just a few cells in an attractor network ("partial" representations with just one or two black dots on the left and right examples in "Attractor 1" and "Attractor 2") is sufficient to restore the full ensemble activity of cells participating in the representation ("complete" patterns shown in the middle of "Attractor 1" and "Attractor 2"). (*C*) Attractor networks are thought to aid in disambiguating similar patterns of input by favoring sharp transitions between network states as inputs are changed gradually, as in the study by Wills, Lever, Cacucci, Burgess, and O'Keefe (2005). It was found that spatial maps in the hippocampus snapped sharply from a "square-environment" representation to a "circle-environment" representation as a recording enclosure was gradually "morphed" from one shape to the other. (Modified with permission from S. Leutgeb, J. K. Leutgeb, Moser, & Moser, Place cells, spatial maps and the population code for memory, *Current Opinion in Neurobiology*, *15*, 738–746, © 2005, Elsevier Ltd.)

be based on LTP- and LTD-like synaptic modifications between the cells that participate in the individual representations and between these ensembles and external signals providing information about the features of the environment or episode for which a representation is generated.

Unfortunately, there is limited direct experimental evidence for LTP and LTD in attractor dynamics. Several recent studies have suggested that the hippocampus *has* attractor properties, however. For example, place cells keep their location of firing after removal of a significant subset of the landmarks that defined the original training environment—for example, when a cue card is removed or the lights are turned off (Muller & Kubie, 1987; O'Keefe & Conway, 1978; Quirk, Muller, & Kubie, 1990). The persistence of the place fields suggests that representations can be activated even under severely degraded input conditions (as schematized in figure 7.8*B*). However, such experiments do not rule out the possibility that firing is controlled by subtle cues that are still present in the deprived version of the environment. In response to this concern, more recent experiments have measured hippocampal place representations during progressive equal-step transformation of the recording environment, using so-called morph boxes. Recording in CA1, Wills and colleagues trained rats in a square and a circular version of a box with flexible walls until place cell representations in the two environments were very different (Wills, Lever, Cacucci, Burgess, & O'Keefe, 2005). The rats were then exposed to several intermediate shapes. A sharp transition from squarelike representations to circlelike representations was observed near the middle between the familiar shapes, as predicted if the network had discrete attractor-based representations corresponding to the trained shapes (figure 7.8*C*). Parallel work by Leutgeb and colleagues showed that the representations are not always discrete (J. Leutgeb et al., 2005). Under conditions where the spatial reference frame is constant, place cells in CA3 and CA1 assimilate gradual or moderate changes in the environment into the preexisting representations. It was observed that stable states can be attained along the entire continuum between two preestablished representations, as long as the spatial environment remains unchanged. Adding the dimension of time, this ability to represent

continua may allow hippocampal networks to encode and retrieve sequential inputs as uninterrupted episodes. The existence of both discrete and continuous representations and their dependence on the exact experience in the environment are consistent with the existence of attractors in the hippocampus, but the attractors must be dynamic, implying a possible role for LTP and LTD in their formation and maintenance.

Where should we begin the study of synaptic plasticity in hippocampal attractor dynamics? Theoretical models have pointed to the neural architecture of CA3 as a good candidate (Marr, 1971; McNaughton & Morris, 1987). The dense and modifiable recurrent circuitry of this system (Amaral & Witter, 1989; Lorente de Nó, 1934) and the sparse firing of the pyramidal cells in this area (Barnes, McNaughton, Mizumori, Leonard, & Lin, 1990; S. Leutgeb, J. K. Leutgeb, Treves, Moser, & Moser, 2004) are properties that would be expected if the system were to form rapid distinguishable representations that could be recalled in the presence of considerable noise. Widespread collaterals interconnect pyramidal cells along nearly the entire length of the CA3 (Ishizuka, Weber, & Amaral, 1990; Li, Somogyi, Ylinen, & Buzsaki, 1994). On average, each pyramidal neuron makes synapses with as many as 4% of the pyramidal cells in the ipsilateral CA3, and more than three-quarters of the excitatory synapses on a CA3 pyramidal cell are from other CA3 pyramidal neurons (Amaral, Ishizuka, & Claiborne, 1990). The recurrent synapses exhibit LTP (Zalutsky & Nicoll, 1990), enabling the formation of an extensive number of interconnected cell groups in the network. In agreement with these ideas, CA3 cells show remarkably persistent and coherent firing after removal of significant parts of the original sensory input (Lee, Yoganarasimha, Rao, & Knierim, 2004; S. Leutgeb et al., 2004; Vazdarjanova & Guzowski, 2004). NMDA-receptor-dependent plasticity in CA3 is necessary for this neural reactivation process as well as the ability to retrieve spatial memories from small subsets of the cues of the original environment (see figure 7.7*B*) (Nakazawa et al., 2002). Formation of discrete representations is apparent in the same network as a nearly complete replacement of the active cell population in CA3 when animals are transferred between enclosures with common features (S. Leutgeb et al., 2004). Thus several studies suggest that CA3 has the predicted properties of an attractor network and that NMDA receptor-dependent long-term plasticity may underlie its functions in encoding and recall of memory.

Summary

There is still some work to do. Demonstrating that LTP plays a role in memory is only a first step to a deeper mechanistic understanding of how the brain achieves information storage and recall. The available data suggest that the question is no longer *whether* LTP is involved in memory, but *how*. A major challenge for future research will be to determine more exactly how LTP and LTD contribute to dynamic representation in the heavily interconnected neural networks of the hippocampus and elsewhere. The evidence for attractors is indirect, and we do not know, for example, what numbers of cells are involved in each representation, whether there are multiple representations, and, if there are, whether and how they overlap and interact. The mechanisms for maintaining and separating discrete representations, as well as the processes by which new information is assimilated into existing network states, are not known. LTP and LTD, as well as more short-term plasticity processes, are likely to play major functions in these processes, but how these functions are implemented in the network remains an enigma.

REFERENCES

Abraham, W. C., & Kairiss, E. W. (1988). Effects of the NMDA antagonist 2AP5 on complex spike discharge by hippocampal pyramidal cells. *Neurosci. Lett.*, *89*(1), 36–42.

Abraham, W. C., Logan, B., Greenwood, J. M., & Dragunow, M. (2002). Induction and experience-dependent consolidation of stable long-term potentiation lasting months in the hippocampus. *J. Neurosci.*, *22*(21), 9626–9634.

Abraham, W. C., & Williams, J. M. (2003). Properties and mechanisms of LTP maintenance. *Neuroscientist*, *9*(6), 463–474.

Amaral, D. G., Ishizuka, N., & Claiborne, B. (1990). Neurons, numbers and the hippocampal network. *Prog. Brain Res.*, *83*, 1–11.

Amaral, D. G., & Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience*, *31*(3), 571–591.

Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1985). Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys. Rev. Lett.*, *55*(14), 1530–1533.

Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1987). Information storage in neural networks with low levels of activity. *Phys. Rev. A*, *35*(5), 2293–2303.

Andersen, P., Sundberg, S. H., Sveen, O., & Wigstrom, H. (1977). Specific long-lasting potentiation of synaptic transmission in hippocampal slices. *Nature*, *266*(5604), 736–737.

Bannerman, D. M., Good, M. A., Butcher, S. P., Ramsay, M., & Morris, R. G. (1995). Distinct components of spatial learning revealed by prior training and NMDA receptor blockade. *Nature*, *378*(6553), 182–186.

Barnes, C. A. (1979). Memory deficits associated with senescence: A neurophysiological and behavioral study in the rat. *J. Comp. Physiol. Psychol.*, *93*(1), 74–104.

Barnes, C. A., McNaughton, B. L., Mizumori, S. J., Leonard, B. W., & Lin, L. H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Prog. Brain Res.*, *83*, 287–300.

Barrionuevo, G., & Brown, T. H. (1983). Associative long-term potentiation in hippocampal slices. *Proc. Natl. Acad. Sci. USA*, *80*(23), 7347–7351.

Bevilaqua, L. R., Medina, J. H., Izquierdo, I., & Cammarota, M. (2005). Memory consolidation induces N-methyl-D-aspartic acid-receptor- and Ca^{2+}/calmodulin-dependent protein kinase II-dependent modifications in alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptor properties. *Neuroscience*, *136*(2), 397–403.

Bliss, T. V., & Collingridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature*, *361*(6407), 31–39.

Bliss, T. V., Errington, M. L., & Lynch, M. A. (1990). Long-term potentiation in the dentate gyrus in vivo is associated with a sustained increase in extracellular glutamate. *Adv. Exp. Med. Biol.*, *268*, 269–278.

Bliss, T. V., Errington, M. L., Lynch, M. A., & Williams, J. H. (1990). Presynaptic mechanisms in hippocampal long-term potentiation. *Cold Spring Harb. Symp. Quant. Biol.*, *55*, 119–129.

Bliss, T. V., & L mo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.*, *232*(2), 331–356.

Bostock, E., Muller, R. U., & Kubie, J. L. (1991). Experience-dependent modifications of hippocampal place cell firing. *Hippocampus*, *1*(2), 193–205.

Browning, M. D., Huganir, R., & Greengard, P. (1985). Protein phosphorylation and neuronal function. *J. Neurochem.*, *45*(1), 11–23.

Brun, V. H., Ytterbo, K., Morris, R. G., Moser, M. B., & Moser, E. I. (2001). Retrograde amnesia for spatial memory induced by NMDA receptor-mediated long-term potentiation. *J. Neurosci.*, *21*(1), 356–362.

Butcher, S. P., Davis, S., & Morris, R. G. (1990). A dose-related impairment of spatial learning by the NMDA receptor antagonist, 2-amino-5-phosphonovalerate (AP5). *Eur. Neuropsychopharmacol.*, *1*(1), 15–20.

Cammarota, M., Bernabeu, R., Levi De Stein, M., Izquierdo, I., & Medina, J. H. (1998). Learning-specific, time-dependent increases in hippocampal Ca^{2+}/calmodulin-dependent protein kinase II activity and AMPA GluR1 subunit immunoreactivity. *Eur. J. Neurosci.*, *10*(8), 2669–2676.

Caramanos, Z., & Shapiro, M. L. (1994). Spatial memory and *N*-methyl-D-aspartate receptor antagonists APV and MK-801: Memory impairments depend on familiarity with the environment, drug dose, and training duration. *Behav. Neurosci.*, *108*(1), 30–43.

Chen, L. Y., Rex, C. S., Casale, M. S., Gall, C. M., & Lynch, G. (2007). Changes in synaptic morphology accompany actin signaling during LTP. *J. Neurosci.*, *27*(20), 5363–5372.

Colledge, M., Snyder, E. M., Crozier, R. A., Soderling, J. A., Jin, Y., Langeberg, L. K., et al. (2003). Ubiquitination regulates PSD-95 degradation and AMPA receptor surface expression. *Neuron*, *40*(3), 595–607.

Cummings, J. A., Mulkey, R. M., Nicoll, R. A., & Malenka, R. C. (1996). Ca^{2+} signaling requirements for long-term depression in the hippocampus. *Neuron*, *16*(4), 825–833.

Dolphin, A. C., Errington, M. L., & Bliss, T. V. (1982). Long-term potentiation of the perforant path in vivo is associated with increased glutamate release. *Nature*, *297*(5866), 496–498.

Douglas, R. M., & Goddard, G. V. (1975). Long-term potentiation of the perforant path-granule cell synapse in the rat hippocampus. *Brain Res.*, *86*(2), 205–215.

Dragoi, G., Harris, K. D., & Buzsaki, G. (2003). Place representation within hippocampal networks is modified by long-term potentiation. *Neuron*, *39*(5), 843–853.

Dudek, S. M., & Bear, M. F. (1992). Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Proc. Natl. Acad. Sci. USA*, *89*(10), 4363–4367.

Eccles, J. C., & McIntyre, A. K. (1953). The effects of disuse and of activity on mammalian spinal reflexes. *J. Physiol.*, *121*(3), 492–516.

Ehlers, M. D. (2000). Reinsertion or degradation of AMPA receptors determined by activity-dependent endocytic sorting. *Neuron*, *28*(2), 511–525.

Ekstrom, A. D., Meltzer, J., McNaughton, B. L., & Barnes, C. A. (2001). NMDA receptor antagonism blocks experience-dependent expansion of hippocampal "place fields." *Neuron*, *31*(4), 631–638.

Frey, U., Krug, M., Reymann, K. G., & Matthies, H. (1988). Anisomycin, an inhibitor of protein synthesis, blocks late phases of LTP phenomena in the hippocampal CA1 region in vitro. *Brain Res.*, *452*(1–2), 57–65.

Frey, U., & Morris, R. G. (1997). Synaptic tagging and long-term potentiation. *Nature*, *385*(6616), 533–536.

Fyhn, M., Hafting, T., Treves, A., Moser, M. B., & Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, *446*(7132), 190–194.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M. B. (2004). Spatial representation in the entorhinal cortex. *Science*, *305*(5688), 1258–1264.

Gruart, A., Munoz, M. D., & Delgado-Garcia, J. M. (2006). Involvement of the CA3-CA1 synapse in the acquisition of associative learning in behaving mice. *J. Neurosci.*, *26*(4), 1077–1087.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806.

Hampson, R. E., Heyser, C. J., & Deadwyler, S. A. (1993). Hippocampal cell firing correlates of delayed-match-to-sample performance in the rat. *Behav. Neurosci.*, *107*(5), 715–739.

Hayashi, Y., Shi, S. H., Esteban, J. A., Piccini, A., Poncer, J. C., & Malinow, R. (2000). Driving AMPA receptors into synapses by LTP and CaMKII: Requirement for GluR1 and PDZ domain interaction. *Science*, *287*(5461), 2262–2267.

Hebb, D. O. (1949). *The organization of behavior*. New York: John Wiley & Sons.

Heynen, A. J., Quinlan, E. M., Bae, D. C., & Bear, M. F. (2000). Bidirectional, activity-dependent regulation of glutamate receptors in the adult hippocampus in vivo. *Neuron*, *28*(2), 527–536.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, *79*(8), 2554–2558.

Ishizuka, N., Weber, J., & Amaral, D. G. (1990). Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J. Comp. Neurol.*, *295*(4), 580–623.

Izquierdo, I., Bevilaqua, L. R., Rossato, J. I., Bonini, J. S., Medina, J. H., & Cammarota, M. (2006). Different molecular cascades in different sites of the brain control memory consolidation. *Trends Neurosci.*, *29*(9), 496–505.

Izquierdo, I., da Cunha, C., Rosat, R., Jerusalinsky, D., Ferreira, M. B., & Medina, J. H. (1992). Neurotransmitter receptors involved in post-training memory processing by the

amygdala, medial septum, and hippocampus of the rat. *Behav. Neural Biol.*, *58*(1), 16–26.

Jerusalinsky, D., Ferreira, M. B., Walz, R., Da Silva, R. C., Bianchin, M., Ruschel, A. C., et al. (1992). Amnesia by post-training infusion of glutamate receptor antagonists into the amygdala, hippocampus, and entorhinal cortex. *Behav. Neural Biol.*, *58*(1), 76–80.

Kelleher, R. J., 3rd, Govindarajan, A., & Tonegawa, S. (2004). Translational regulatory mechanisms in persistent forms of synaptic plasticity. *Neuron*, *44*(1), 59–73.

Kentros, C., Hargreaves, E., Hawkins, R. D., Kandel, E. R., Shapiro, M., & Muller, R. V. (1998). Abolition of long-term stability of new hippocampal place cell maps by NMDA receptor blockade. *Science*, *280*(5372), 2121–2126.

Krug, M., Lossner, B., & Ott, T. (1984). Anisomycin blocks the late phase of long-term potentiation in the dentate gyrus of freely moving rats. *Brain Res. Bull.*, *13*(1), 39–42.

Lee, I., Yoganarasimha, D., Rao, G., & Knierim, J. J. (2004). Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. *Nature*, *430*(6998), 456–459.

Leutgeb, J. K., Leutgeb, S., Treves, A., Meyer, R., Barnes, C. A., McNaughton, B. L., et al. (2005). Progressive transformation of hippocampal neuronal representations in "morphed" environments. *Neuron*, *48*(2), 345–358.

Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., & Moser, M. B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, *309*(5734), 619–623.

Leutgeb, S., Leutgeb, J. K., Moser, M. B., & Moser, E. I. (2005). Place cells, spatial maps and the population code for memory. *Curr. Opin. Neurobiol.*, *15*(6), 738–746.

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M. B., & Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science*, *305*(5688), 1295–1298.

Levy, W. B., & Steward, O. (1979). Synapses as associative memory elements in the hippocampal formation. *Brain Res.*, *175*(2), 233–245.

Li, X. G., Somogyi, P., Ylinen, A., & Buzsaki, G. (1994). The hippocampal CA3 network: An in vivo intracellular labeling study. *J. Comp. Neurol.*, *339*(2), 181–208.

Lisman, J. E. (1985). A mechanism for memory storage insensitive to molecular turnover: A bistable autophosphorylating kinase. *Proc. Natl. Acad. Sci. USA*, *82*(9), 3055–3057.

Lorente de Nó, R. (1934). Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *J. Psychol. Neurol.*, *46*, 113–177.

Madison, D. V., Malenka, R. C., & Nicoll, R. A. (1991). Mechanisms underlying long-term potentiation of synaptic transmission. *Annu. Rev. Neurosci.*, *14*, 379–397.

Malenka, R. C., & Bear, M. F. (2004). LTP and LTD: An embarrassment of riches. *Neuron*, *44*(1), 5–21.

Malenka, R. C., & Nicoll, R. A. (1999). Long-term potentiation—A decade of progress? *Science*, *285*(5435), 1870–1874.

Malinow, R., Mainen, Z. F., & Hayashi, Y. (2000). LTP mechanisms: From silence to four-lane traffic. *Curr. Opin. Neurobiol.*, *10*(3), 352–357.

Malinow, R., & Malenka, R. C. (2002). AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.*, *25*, 103–126.

Mansuy, I. M., Winder, D. G., Moallem, T. M., Osman, M., Mayford, M., Hawkins, R. D., et al. (1998). Inducible and reversible gene expression with the rtTA system for the study of memory. *Neuron*, *21*(2), 257–265.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.*, *24*, 897–931.

Markus, E. J., Qin, Y. L., Leonard, B., Skaggs, W. E., McNaughton, B. L., & Barnes, C. A. (1995). Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *J. Neurosci.*, *15*(11), 7079–7094.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *262*(841), 23–81.

Martin, S. J., Grimwood, P. D., & Morris, R. G. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annu. Rev. Neurosci.*, *23*, 649–711.

Matsuo, N., Reijmers, L., & Mayford, M. (2008). Spine-type-specific recruitment of newly synthesized AMPA receptors with learning. *Science*, *319*(5866), 1104–1107.

Mayer, M. L., Westbrook, G. L., & Guthrie, P. B. (1984). Voltage-dependent block by Mg^{2+} of NMDA responses in spinal cord neurones. *Nature*, *309*(5965), 261–263.

Mayford, M., & Kandel, E. R. (1999). Genetic approaches to memory storage. *Trends Genet.*, *15*(11), 463–470.

McHugh, T. J., Blum, K. I., Tsien, J. Z., Tonegawa, S., & Wilson, M. A. (1996). Impaired hippocampal representation of space in CA1-specific NMDAR1 knockout mice. *Cell*, *87*(7), 1339–1349.

McKernan, M. G., & Shinnick-Gallagher, P. (1997). Fear conditioning induces a lasting potentiation of synaptic currents in vitro. *Nature*, *390*(6660), 607–611.

McNaughton, B. L., Barnes, C. A., Rao, G., Baldwin, J., & Rasmussen, M. (1986). Long-term enhancement of hippocampal synaptic transmission and the acquisition of spatial information. *J. Neurosci.*, *6*(2), 563–571.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the "cognitive map." *Nat. Rev. Neurosci.*, *7*(8), 663–678.

McNaughton, B. L., Douglas, R. M., & Goddard, G. V. (1978). Synaptic enhancement in fascia dentata: Cooperativity among coactive afferents. *Brain Res.*, *157*(2), 277–293.

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.*, *10*(10), 408–414.

Mehta, M. R., Barnes, C. A., & McNaughton, B. L. (1997). Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc. Natl. Acad. Sci. USA*, *94*(16), 8918–8921.

Mehta, M. R., Quirk, M. C., & Wilson, M. A. (2000). Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron*, *25*(3), 707–715.

Moita, M. A., Rosis, S., Zhou, Y., LeDoux, J. E., & Blair, H. T. (2004). Putting fear in its place: Remapping of hippocampal place cells during fear conditioning. *J. Neurosci.*, *24*(31), 7015–7023.

Monfils, M. H., & Teskey, G. C. (2004). Skilled-learning-induced potentiation in rat sensorimotor cortex: A transient form of behavioural long-term potentiation. *Neuroscience*, *125*(2), 329–336.

Morris, R. G., Anderson, E., Lynch, G. S., & Baudry, M. (1986). Selective impairment of learning and blockade of long-term potentiation by an *N*-methyl-D-aspartate receptor antagonist, AP5. *Nature*, *319*(6056), 774–776.

Moser, E. I., Krobert, K. A., Moser, M. B., & Morris, R. G. (1998). Impaired spatial learning after saturation of long-term potentiation. *Science*, *281*(5385), 2038–2042.

Mulkey, R. M., Endo, S., Shenolikar, S., & Malenka, R. C. (1994). Involvement of a calcineurin/inhibitor-1 phosphatase cascade in hippocampal long-term depression. *Nature*, *369*(6480), 486–488.

Mulkey, R. M., Herron, C. E., & Malenka, R. C. (1993). An essential role for protein phosphatases in hippocampal long-term depression. *Science*, *261*(5124), 1051–1055.

Mulkey, R. M., & Malenka, R. C. (1992). Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus. *Neuron*, *9*(5), 967–975.

Muller, R. U., & Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J. Neurosci.*, *7*(7), 1951–1968.

Nagerl, U. V., Eberhorn, N., Cambridge, S. B., & Bonhoeffer, T. (2004). Bidirectional activity-dependent morphological plasticity in hippocampal neurons. *Neuron*, *44*(5), 759–767.

Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., et al. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science*, *297*(5579), 211–218.

Neves, G., Cooke, S. F., & Bliss, T. V. (2008). Synaptic plasticity, memory and the hippocampus: A neural network approach to causality. *Nat. Rev. Neurosci.*, *9*(1), 65–75.

Nowak, L., Bregestovski, P., Ascher, P., Herbet, A., & Prochiantz, A. (1984). Magnesium gates glutamate-activated channels in mouse central neurones. *Nature*, *307*(5950), 462–465.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Exp. Neurol.*, *51*(1), 78–109.

O'Keefe, J., & Conway, D. H. (1978). Hippocampal place units in the freely moving rat: Why they fire where they fire. *Exp. Brain Res.*, *31*(4), 573–590.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.*, *34*(1), 171–175.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford, UK: Clarendon Press.

Paratcha, G., Furman, M., Bevilaqua, L., Cammarota, M., Vianna, M., de Stein, M. L., et al. (2000). Involvement of hippocampal PKCbetaI isoform in the early phase of memory formation of an inhibitory avoidance learning. *Brain Res.*, *855*(2), 199–205.

Pastalkova, E., Serrano, P., Pinkhasova, D., Wallace, E., Fenton, A. A., & Sacktor, T. C. (2006). Storage of spatial information by the maintenance mechanism of LTP. *Science*, *313*(5790), 1141–1144.

Power, J. M., Thompson, L. T., Moyer, J. R., Jr., & Disterhoft, J. F. (1997). Enhanced synaptic transmission in CA1 hippocampus after eyeblink conditioning. *J. Neurophysiol.*, *78*(2), 1184–1187.

Quirk, G. J., Muller, R. U., & Kubie, J. L. (1990). The firing of hippocampal place cells in the dark depends on the rat's recent experience. *J. Neurosci.*, *10*(6), 2008–2017.

Rioult-Pedotti, M. S., Donoghue, J. P., & Dunaevsky, A. (2007). Plasticity of the synaptic modification range. *J. Neurophysiol.*, *98*(6), 3688–3695.

Rioult-Pedotti, M. S., Friedman, D., & Donoghue, J. P. (2000). Learning-induced LTP in neocortex. *Science*, *290*(5491), 533–536.

Rioult-Pedotti, M. S., Friedman, D., Hess, G., & Donoghue, J. P. (1998). Strengthening of horizontal cortical connections following skill learning. *Nat. Neurosci.*, *1*(3), 230–234.

Rogan, M. T., Staubli, U. V., & LeDoux, J. E. (1997). Fear conditioning induces associative long-term potentiation in the amygdala. *Nature*, *390*(6660), 604–607.

Rumpel, S., LeDoux, J., Zador, A., & Malinow, R. (2005). Postsynaptic receptor trafficking underlying a form of associative learning. *Science*, *308*(5718), 83–88.

Sacchetti, B., Lorenzini, C. A., Baldi, E., Bucherelli, C., Roberto, M., Tassoni, G., et al. (2001). Long-lasting hippocampal potentiation and contextual memory consolidation. *Eur. J. Neurosci.*, *13*(12), 2291–2298.

Sacchetti, B., Lorenzini, C. A., Baldi, E., Bucherelli, C., Roberto, M., Tassoni, G., et al. (2002). Time-dependent inhibition of hippocampal LTP in vitro following contextual fear conditioning in the rat. *Eur. J. Neurosci.*, *15*(1), 143–150.

Sajikumar, S., & Frey, J. U. (2004). Late-associativity, synaptic tagging, and the role of dopamine during LTP and LTD. *Neurobiol. Learn. Mem.*, *82*(1), 12–25.

Saucier, D., & Cain, D. P. (1995). Spatial learning without NMDA receptor-dependent long-term potentiation. *Nature*, *378*(6553), 186–189.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry*, *20*(1), 11–21.

Shi, S. H., Hayashi, Y., Petralia, R. S., Zaman, S. H., Wenthold, R. J., Svoboda, K., et al. (1999). Rapid spine delivery and redistribution of AMPA receptors after synaptic NMDA receptor activation. *Science*, *284*(5421), 1811–1816.

Shors, T. J., & Matzel, L. D. (1997). Long-term potentiation: What's learning got to do with it? *Behav. Brain. Sci.*, *20*(4), 597–614; discussion, 614–655.

Shukla, K., Kim, J., Blundell, J., & Powell, C. M. (2007). Learning-induced glutamate receptor phosphorylation resembles that induced by long term potentiation. *J. Biol. Chem.*, *282*(25), 18100–18107.

Soderling, T. R., & Derkach, V. A. (2000). Postsynaptic protein phosphorylation and LTP. *Trends Neurosci.*, *23*(2), 75–80.

Stanton, P. K., & Sarvey, J. M. (1984). Blockade of long-term potentiation in rat hippocampal CA1 region by inhibitors of protein synthesis. *J. Neurosci.*, *4*(12), 3080–3088.

Steele, R. J., & Morris, R. G. (1999). Delay-dependent impairment of a matching-to-place task with chronic and intrahippocampal infusion of the NMDA-antagonist D-AP5. *Hippocampus*, *9*(2), 118–136.

Stevens, C. F. (1998). A million dollar question: Does LTP = memory? *Neuron*, *20*(1), 1–2.

Thompson, L. T., & Best, P. J. (1990). Long-term stability of the place-field activity of single units recorded from the dorsal hippocampus of freely behaving rats. *Brain Res.*, *509*(2), 299–308.

Tsien, J. Z., Huerta, P. T., & Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell*, *87*(7), 1327–1338.

Vazdarjanova, A., & Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *J. Neurosci.*, *24*(29), 6489–6496.

Whitlock, J. R., Heynen, A. J., Shuler, M. G., & Bear, M. F. (2006). Learning induces long-term potentiation in the hippocampus. *Science*, *313*(5790), 1093–1097.

Wills, T. J., Lever, C., Cacucci, F., Burgess, N., & O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment. *Science*, *308*(5723), 873–876.

Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, *261*(5124), 1055–1058.

Wolfman, C., Fin, C., Dias, M., Bianchin, M., Da Silva, R. C., Schmitz, P. K., et al. (1994). Intrahippocampal or intra-amygdala infusion of KN62, a specific inhibitor of calcium/calmodulin-dependent protein kinase II, causes retrograde amnesia in the rat. *Behav. Neural Biol.*, *61*(3), 203–205.

Wood, E. R., Dudchenko, P. A., & Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, *397*(6720), 613–616.

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., & Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, *27*(3), 623–633.

Young, B. J., Fox, G. D., & Eichenbaum, H. (1994). Correlates of hippocampal complex–spike cell activity in rats performing a nonspatial radial maze task. *J. Neurosci.*, *14*(11, Pt. 1), 6553–6563.

Yuste, R., & Bonhoeffer, T. (2001). Morphological changes in dendritic spines associated with long-term synaptic plasticity. *Annu. Rev. Neurosci.*, *24*, 1071–1089.

Zalutsky, R. A., & Nicoll, R. A. (1990). Comparison of two forms of long-term potentiation in single hippocampal neurons. *Science*, *248*(4963), 1619–1624.

Zhou, Q., Homma, K. J., & Poo, M. M. (2004). Shrinkage of dendritic spines associated with long-term depression of hippocampal synapses. *Neuron*, *44*(5), 749–757.

8 Visual Cortical Plasticity and Perceptual Learning

WU LI AND CHARLES D. GILBERT

ABSTRACT Plasticity is an integral property of a functioning brain throughout life. In the visual system, cortical plasticity is engaged for encoding the geometric regularities of the visual environment early in life, as well as for functionally adaptive changes in response to lesions and neurodegenerative diseases. In addition to the pliability during postnatal maturation and during the restoration of disrupted functions, the visual system also maintains remarkable plasticity for encoding the specific shapes of figures to which we become familiar. This is known as perceptual learning, and it is important for rapid recognition of the learned shapes in complex environments and for enhanced sensitivity to delicate nuances of the learned stimulus features. Moreover, the visual system also exhibits fast functional switching capabilities, whereby response properties of neurons are dynamically adjusted by top-down influences for efficient processing of behaviorally relevant stimuli. The dynamic nature of neuronal responses is tightly coupled with the long-term plasticity seen in perceptual learning, as repeated performing of the same perceptual task, and therefore, repetitive invoking of top-down influences specific to the task, can potentiate the dynamic changes useful for solving the perceptual tasks, leading to encoding and retrieving of the implicit memory formed during perceptual learning.

Our brain needs to constantly adapt to the environment and to assimilate knowledge about the external world by maintaining a certain degree of functional and architectural malleability. This notion has been appreciated for centuries. The idea that our perceptual and cognitive functions can be shaped by an individual's experience was originally expressed by philosophers such as John Locke, who asserted that the human mind at birth is like a blank slate, and that all ideas and knowledge are derived from individual's experiences (Locke, 1689/1995). The earliest psychological inference and definition of cortical plasticity were made by William James (1890/1950), who compared the formation of habits and skills to the plastic changes of materials, and attributed behavioral changes to the plasticity of the brain. One of the most influential speculations about the neuronal substrates of cortical plasticity was vividly drawn by Santiago Ramón y Cajal (1911), who proposed that changes in connections between neurons are responsible for our ability to learn (see review, Jones, 1994). As for the rule governing wiring and rewiring between neurons, Donald Hebb theoretically postulated that neurons are wired together if they fire together (Hebb, 1949). This Hebbian rule of synaptic plasticity has been widely adopted into physiological, psychophysical, and computational studies of learning and memory. At the system and behavior levels, Jerzy Konorski (1948) distinguished plasticity from excitability as an independent property of the brain whereby "certain permanent functional transformations arise in particular systems of neurons as the result of appropriate stimuli." On top of these earlier insightful reasoning and speculations, the last half century has witnessed the advances of our understanding of cortical plasticity in various respects, from different perspectives, and using a variety of approaches. This chapter focuses on the cortical plasticity in the visual system.

Processing of visual information in the brain is distributed among more than 30 cortical areas (Van Essen, Anderson, & Felleman, 1992). These functionally specialized and hierarchically organized areas are interwired by feedforward and feedback connections, forming partially segregated modules and pathways for processing different attributes of visual stimuli. On the one hand, this specific connectivity has been genetically determined or innately hardwired for mediating both stimulus-driven bottom-up process and behavior-driven top-down influences. On the other hand, accumulated evidence has revealed that visual experience can modify the preexisting functionality and connectivity of the visual system throughout life.

Plasticity in postnatal development

EARLY DEVELOPMENT The maturation process of the visual system continues well into postnatal periods in terms of both circuitry and functionality. The neural circuitry within a cortical area, such as the primary visual cortex (area V1), comprises two types of connections (for reviews see Gilbert, 1983; Callaway, 1998). The vertical connections, which link neurons across different cortical layers that represent the same visual field location, are responsible for processing local simple stimulus attributes. The horizontal connections, which extend parallel to the cortical surface

WU LI Beijing Normal University, Beijing, China
CHARLES D. GILBERT The Rockefeller University, New York, New York

and tend to link cells with nonoverlapping receptive fields (RFs), are involved in integrating information over a large area in the visual field (for review see Gilbert, 1992). Anatomical studies have shown that the vertical connections in human V1 develop prenatally, but the development of the horizontal connections is much later and is asynchronous within different cortical layers (Callaway & Katz, 1990; Burkhalter, Bernardo, & Charles, 1993). In the middle (the input) layers, the horizontal connections propagate rapidly only after birth and become more adultlike in about two months. The horizontal connections in the superficial (the output) layers develop at an even later age, emerging after birth and reaching mature form after more than one year. The anisotropy of development and maturation of local circuits in V1 suggests that different visual functions emerge at different stages of development and that postnatal experience could be important for the maturation of cortical circuitry and visual functions. Recent studies have provided compelling evidence that early visual experience can influence the maturation of neural circuitry by shaping axonal and dendritic structures, regulating synapse formation and elimination, and altering synaptic transmission (for review see Fox & Wong, 2005). These experience-dependent changes during postnatal development can be observed at different levels of the visual pathway, including the visual cortex, the lateral geniculate nucleus, and even the retina.

The susceptibility of functional architecture of visual cortex to anomalous visual experience was first discovered by Hubel and Wiesel (Wiesel & Hubel, 1963; Hubel & Wiesel, 1970, 1977). In the normal developmental condition, the visual input to area V1 from the two eyes is balanced. This balance can be disrupted by depriving an animal of the visual input from one eye within a period of several months after birth, resulting in an enlargement in representation of the normal eye and shrinkage in the representation of the deprived eye. The time window within which a brain function is highly susceptible to experience-dependent modifications is referred to as the critical period. Depending on different brain functions and different species, the onset and closure of the critical period can be different. For example, in humans the critical period for susceptibility of stereopsis begins soon after birth and extends into childhood for at least 4–5 years (Fawcett, Wang, & Birch, 2005). Since the heightened plasticity is maintained only within a finite period of postnatal development, it leads to the conjecture that the response properties of neurons and the functional architecture of the cortex become fixed in adulthood. As we will see in the other sections of this chapter, this idea has been challenged in the last couple of decades.

Late Maturation An extreme case of late maturation of some visual functions is seen in the perceptual ability to link discrete contour elements into a global coherent visual contour within a complex background (for a demonstration of visual contours, see figure 8.1). This process, known as contour integration, is an important intermediate step in object recognition. According to the Gestalt rule of "good continuation," our visual system has built-in apparatus to link contour elements that are arranged along smooth and continuous paths (Wertheimer, 1923). Recent studies have suggested that the long-range horizontal connections in V1, which link neurons with similar preference for contour orientations, are ideally suited for mediating contour integration, both in terms of its orientation specificity and its spatial extent (W. Li & Gilbert, 2002; Stettler, Das, Bennett, & Gilbert, 2002; W. Li, Piech, & Gilbert, 2006). Moreover, this hardwired connectivity ecologically coincides with the geometries and regularities of natural scene images that are rich in collinear and cocircular contours (Geisler, Perry, Super, & Gallogly, 2001; Sigman, Cecchi, Gilbert, & Magnasco, 2001), suggesting an evolutionary and developmental impact on the formation and maturation of the circuitry. As mentioned previously, the horizontal connections in V1 do not become adultlike until late infancy (Burkhalter et al., 1993). Moreover, it has been shown that the ability of children to detect visual contours camouflaged

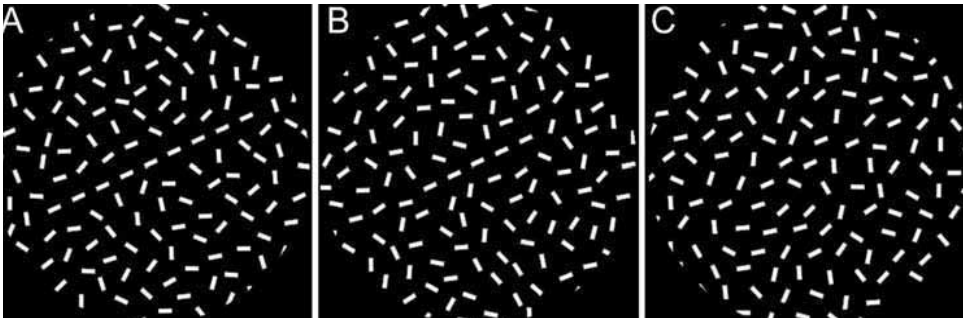


Figure 8.1 Contour integration. Within a complex background, those discrete line segments following the Gestalt law of continuity are easily grouped together, forming a visual contour. A contour consisting of more collinear lines is more salient than a shorter one (compare *A* with *B*); and the same array of collinear lines appears less salient when they are spaced further apart (compare *B* with *C*). (From W. Li, Piech, & Gilbert, 2008.)

in a complex background improves with age and does not approximate the adult's level until adolescence (Kovacs, Kozma, Feher, & Benedek, 1999). Surface segmentation, another important intermediate level visual function, also matures at a late age comparable to contour integration (Sireteanu & Rieth, 1992). Similar to contour integration, the process of partitioning visual images into segregated surfaces relies heavily on integration of information across a large visual field area. The late maturation of contour-integration and surface-segmentation capabilities suggests that natural scene geometries and regularities continue to shape neural circuitry as well as response properties of visual neurons during a very long period of time after birth.

Plasticity in response to lesions

The closure of critical periods does not necessarily mean that the neural connections and circuits in the adult brain have been completely fixed. Abnormal experiences like injuries and neurodegenerative diseases during adulthood can also trigger marked plastic reactions in the central nervous system.

Lesion Experiments Pronounced changes were first reported in the spinal cord of adult animals after an injury to the periphery nerves (Devor & Wall, 1978, 1981). Subsequently, striking reorganization in adult primary sensory cortices has also been widely demonstrated, including the somatosensory cortex in response to deafferentation of sensory input from a skin area (Rasmusson, 1982; Merzenich et al., 1983a, 1983b, 1984; Calford & Tweedale, 1988; Pons et al., 1991; Weiss, Miltner, Liepert, Meissner, & Taub, 2004), the primary auditory cortex in response to restricted cochlear lesions (Robertson & Irvine, 1989; Rajan, Irvine, Wise, & Heil, 1993), and the primary visual cortex in response to lesions on the retina (Kaas et al., 1990; Heinen & Skavenski, 1991; Chino, Kaas, Smith, Langston, & Cheng, 1992; Gilbert & Wiesel, 1992; Schmid, Rosa, Calford, & Ambler, 1996; Eysel et al., 1999; Calford et al., 2000; Giannikopoulos & Eysel, 2006). All these lesion-induced plastic changes have comparable effects in the relevant cortical regions: the cortical territory devoted to representing the deafferented region on the sensory surface (the skin, the cochlea, or the retina) becomes responsive to adjacent sensory surfaces spared from the lesion, a process referred to as cortical reorganization.

Here we use retinal lesions as an example. The retina is mapped point-by-point onto the primary visual cortex, generating a two-dimensional topographic map called the retinotopic map. A restricted lesion on the retina destroys the photoreceptors within a small area (figure 8.2*A*). This retinal scotoma cuts off visual input to the corresponding retinotopic region in V1, known as the lesion projection zone (LPZ), and silences neurons within that cortical region (figure 8.2*B*). After the lesion, continuous plastic changes in V1 have been observed within a period of time ranging from minutes to months (for example, see Gilbert & Wiesel, 1992). Within minutes after the lesion, a remarkable increase in RF sizes occurs for V1 neurons whose RFs are located near the boundary of the retinal scotoma. A couple of months after the retinal injury, the size of the LPZ dramatically shrinks (figure 8.2*C*): neurons within the original cortical LPZ regain responsiveness by shifting their RFs outside the retinal scotoma. This plastic change is not simply a consequence of a rearrangement of thalamocortical afferents; but rather, it is cortically mediated through the long-range horizontal connections intrinsic to V1 (Gilbert & Wiesel, 1992; Darian-Smith & Gilbert, 1995; Calford, Wright, Metha, & Taglianetti, 2003).

Even for the intact visual system in adults, a dramatic change in visual experiences by itself can cause a large-scale functional reorganization of the visual cortex. V1 neurons in a cerebral hemisphere are driven by inputs from the contralateral visual field. After monkeys wore special spectacles for several months to reverse their left and right visual field, some V1 cells begin to respond to stimuli presented in both hemifields (Sugita, 1996).

Neurodegenerative Diseases Similar to the retinal lesion experiments, macular degeneration (MD) has also been

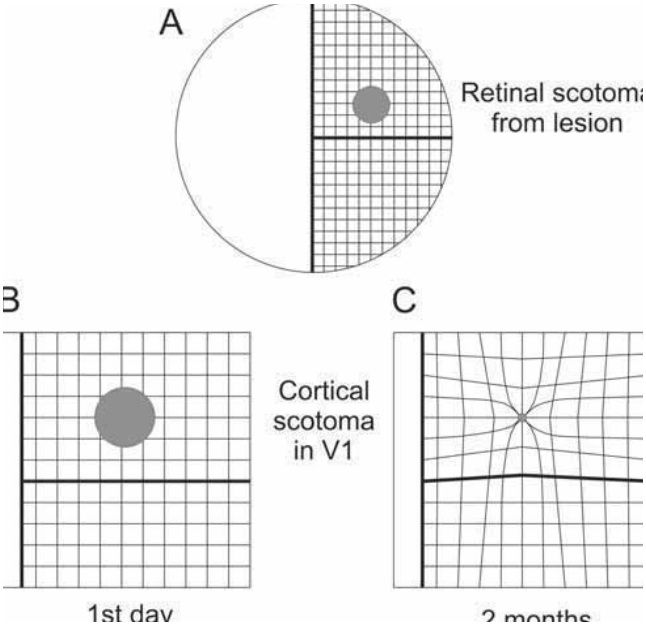


Figure 8.2 Reorganization of V1 in response to retinal lesion. A retinal scotoma produced by focal laser lesion (*A*, the small gray area) creates a silent region in V1 (*B*, the gray area). During recovery (*C*), neurons within the cortical scotoma regain responsiveness to visual input from the retinal area surrounding the laser-induced scotoma. (Adapted from Gilbert, 1992.)

reported to cause large-scale reorganization of the visual cortex, including area V1 (Baker, Peli, Knouf, & Kanwisher, 2005). Macular degeneration is the leading cause of vision loss in old people. It results from a deterioration of the macula, the central area of the retina that offers the highest spatial resolutions in visual processing. Damage to the macula results in deprivation of visual input to the V1 regions representing the central visual field. It has been shown that in MD patients the cortical regions that normally respond to central visual stimuli are strongly activated by peripheral stimuli (Baker et al., 2005; but see Masuda, Dumoulin, Nakadomari, & Wandell, 2008, which argued that the responses seen in the cortical scotoma of MD patients could largely result from top-down influences mediated by feedback connections from higher cortical areas).

In addition to retinal damages, an fMRI study has shown significant cortical reorganization in V1 of an adult patient who suffered from a stroke that destroyed the input fibers to V1 corresponding to the upper-left visual field (Dilks, Serences, Rosenau, Yantis, & McCloskey, 2007). Behavioral tests showed that this patient perceived stimuli located in the intact lower-left visual field as being elongated upward into the blind upper-left visual field. Correspondingly, fMRI experiments confirmed that the deprived V1 region originally representing the blind visual-field area had reorganized to be responsive to the intact visual-field area. This study provides convincing evidence in support of cortical reorganization and its impact on visual perception during recovery of lost visual functions.

Although visual cortex is specialized in processing visual information, in extreme cases like complete sight loss, visual cortical areas can also be recruited to process other sensory information such as enhancing auditory and tactile processing, or even recruited to process higher-order cognitive tasks (for reviews see Burton, 2003; Pascual-Leone, Amedi, Fregni, & Merabet, 2005).

Perceptual learning

Psychophysics Learning-induced modification of perception had already been noticed in the 19th century by Herrmann von Helmholtz, who made an incisive observation that "the judgment of the senses may be modified by experience and by training derived under various circumstances, and may be adapted to the new conditions. Thus, persons may learn in some measure to utilize details of the sensation which otherwise would escape notice and not contribute to obtaining any idea of the object" (Helmholtz, 1866, page 5). Studies within the last couple of decades have identified two distinct classes of long-term learning and memory (for review see Squire, Stark, & Clark, 2004): the explicit or declarative form of memory mediated by a unified memory system—the medial temporal lobe; and the implicit or nondeclarative form of memory distributed in different cortical areas or brain structures. Perceptual learning falls within the category of implicit memory. It is the unconscious acquisition of improved ability with practice in simple perceptual tasks, as demonstrated in many visual tasks, such as discrimination of spatial resolution (McKee & Westheimer, 1978; Poggio, Fahle, & Edelman, 1992; Saarinen & Levi, 1995; Crist, Kapadia, Westheimer, & Gilbert, 1997), orientation (Vogels & Orban, 1985; Shiu & Pashler, 1992; Schoups, Vogels, Qian, & Orban, 2001), direction of motion (Ball & Sekuler, 1982), depth (Fendick & Westheimer, 1983; Ramachandran & Braddick, 1973), texture (Karni & Sagi, 1991), the waveform of sinusoidal stimuli (Fiorentini & Berardi, 1980), and contrast (Yu, Klein, & Levi, 2004). Perceptual training leads to a substantial decrease in the threshold for discriminating subtle changes in the trained stimulus attributes; it also facilitates detection of familiar shapes embedded in an array of similar distracters (Wang, Cavanagh, & Green, 1994; Ellison & Walsh, 1998; Sigman & Gilbert, 2000; Sireteanu & Rettenbach, 2000).

An indication of the possible cortical loci of perceptual learning comes from a large number of psychophysical observations that the learning effect is usually specific to the trained visual field location and to simple stimulus attributes (for review see Gilbert, Sigman, & Crist, 2001). There is little transfer or interference of learning between the trained and untrained visual-field locations, and between different stimuli. For example, training human subjects to discriminate the differences in orientation of an orientated stimulus improves discrimination performance that is only restricted to the trained location and the trained orientation (Schoups, Vogels, & Orban, 1995). Specificity of perceptual learning suggests that learning-induced changes involve early visual cortex like V1, where the visual field is topographically mapped at a fine scale and neurons are selective for simple stimulus attributes like orientation. However, some studies argue that the improved performance could simply be a consequence of retuning the readout connections between the decision stage and the visual cortex, namely, a refinement of the decision criteria (Dosher & Lu, 1998; R. Li, Levi, & Klein, 2004; Yu et al., 2004). A dichotomy of learning mechanisms has also been proposed, whereby some training mainly enhances sensory processing while the other mainly improves decision making (Adini, Wilkonsky, Haspel, Tsodyks, & Sagi, 2004). Despite a lack of general consensus about the cortical loci where the plastic changes take place, physiological and imaging studies have shed light on neural mechanisms of perceptual learning.

Cortical Recruitment Cortical plasticity associated with perceptual learning was first demonstrated in the

somatosensory and auditory systems. The observed changes are analogous to the cortical reorganization observed in the primary somatosensory and auditory cortices in response to peripheral lesions. For example, training monkeys to perform a tactile frequency discrimination task using a restricted skin area induces remarkable reorganization of the primary somatosensory cortex, leading to a significant increase in the size and complexity of the territory representing the trained skin area (Recanzone, Merzenich, & Jenkins, 1992; Recanzone, Merzenich, Jenkins, Grajski, & Dinse, 1992). Similarly, training on an acoustic frequency discrimination task dramatically increases the cortical territory representing the trained frequencies in the primary auditory cortex (Recanzone, Schreiner, & Merzenich, 1993). This mechanism has been referred to as cortical recruitment, whereby a larger cortical region and thus a greater number of neurons are recruited to encode the trained stimuli. Nonetheless, it is still a matter of debate whether the cortical recruitment is directly responsible for the improved discrimination ability, as the recruitment seems unnecessary for enhanced performance on acoustic frequency discrimination (Brown, Irvine, & Park, 2004). Moreover, overrepresentation of the familiar frequencies in the auditory cortex could even be detrimental to discrimination of the overrepresented frequencies (Han, Kover, Insanally, Semerdjian, & Bao, 2007).

In the visual system, an fMRI study has shown that practicing a coherent-motion detection task, in which a small proportion of randomly positioned dots move in the same direction among randomly moving dots, causes a significant enlargement of the cortical territory representing the trained stimulus in area MT, a cortical area involved in motion processing (Vaina, Belliveau, Roziers, & Zeffiro, 1998). However, cortical recruitment associated with perceptual training has never been documented so far in early visual areas (for an attempt to search for such a change in V1, see Crist, Li, & Gilbert, 2001). The lack of transfer or interference of learning across visual field locations and between visual stimuli also argues against cortical recruitment as an effective mechanism of visual perceptual learning, because recruiting by "robbing" adjacent cortical regions would inevitably interfere with processing of other stimuli. However, studies in search of the neural basis of perceptual learning have shown some other cortical changes that can better account for the observed learning effects.

Neuronal Mechanisms The visual stimuli and tasks used for studies of perceptual learning can be roughly put into two categories: visual discrimination and visual detection or identification. In discrimination tasks, observers need to discriminate a subtle change in stimulus with respect to a reference dimension or attribute, such as an orientation discrimination task, to judge whether a line is slightly tilted to the left or to the right with respect to the vertical. In detection tasks, either a target presented alone near its contrast detection threshold or a target embedded in a background of noise or distracters needs to be identified as present or absent. Instead of getting more neurons involved by recruiting, other potential mechanisms to improve performance on these tasks are to increase neuronal selectivity for the stimulus attribute that is relevant to the discrimination task, or to enhance signal-to-noise ratio by selectively boosting neuronal responsiveness to the familiar target, or to achieve automatization and accelerated processing speed by shifting cortical representation of the learned stimulus from higher to lower cortical areas. Neural correlates in all these respects have been found in visual cortical areas, including V1—the first stage of cortical visual processing.

Increased neuronal selectivity in discrimination learning Simple discrimination tasks, such as orientation discrimination, only involve processing of a basic stimulus attribute. It has been shown that training monkeys on orientation discrimination selectively sharpens orientation-tuning functions of those V1 neurons whose RFs are at the trained visual field location and whose preferred orientations are close to the trained orientation (Schoups et al., 2001; but see Ghose, Yang, & Maunsell, 2002). Similar and stronger effects have also been observed in area V4, an intermediate stage in the visual pathway responsible for object recognition (Yang & Maunsell, 2004; Raiguel, Vogels, Mysore, & Orban, 2006). The theoretical interpretations of these observations are mixed. Intuitively, a sharpening of the orientation-tuning curve around the trained orientation would result in an increase in neuronal selectivity for the trained orientation, which would in turn benefit the discrimination task. This idea is supported by a computational study (Teich & Qian, 2003). Conversely, a modeling study argues that a sharpening of original tuning curves actually causes a general loss of information content conveyed by neuronal responses (Series, Latham, & Pouget, 2004).

Unlike discrimination of a simple stimulus attribute, some discrimination tasks require lateral integration of contextual information. The visual percept of a stimulus, as well as responses of visual neurons to the stimulus, can be modified by the global stimulus context within which the stimulus is displayed (for reviews see Gilbert, 1998; Albright & Stoner, 2002; Allman, Miezin, & McGuinness, 1985). This phenomenon, known as contextual modulation, takes place throughout visual cortical areas along the visual pathways, representing a general lateral integrative mechanism of visual processing. Contextual interactions seen in V1 indicate that V1 neurons are selective for more complex features in visual scenes in addition to simple stimulus attributes like contour orientation. It has been shown that extensive

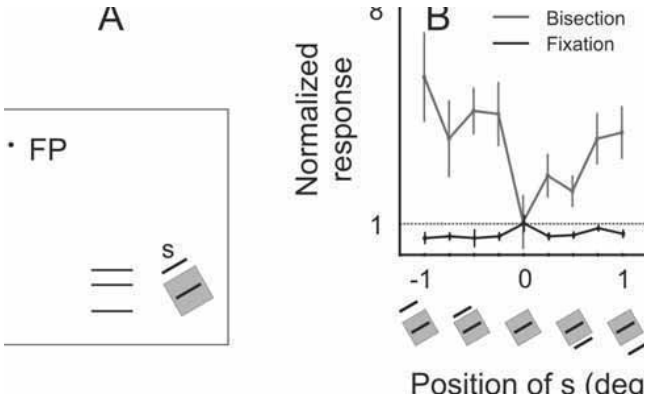


FIGURE 8.3 Perceptual learning modifies contextual influences in V1. (*A*) The stimulus paradigm. The three horizontal parallel lines indicate the task stimulus. Monkeys were trained to determine whether the middle line was closer to the upper or the lower flanker. After training monkeys on this bisection discrimination task, responses of single V1 neurons to another stimulus, the test stimulus, were recorded when the animal either performed the trained bisection task, or simply maintained its fixation at the fixation point (FP). The test stimulus consisted of two lines, an optimally oriented line fixed in the center of the receptive field (denoted by the gray square), and a second parallel line (indicated by "s") placed at different locations on either side of the RF (see the cartoons at the bottom of *B*). (*B*) The normalized responses of a typical V1 cell to the test stimulus as a function of the position of line "s." When the animal was performing the simple fixation task, placing "s" on either side of the RF slightly suppressed neuronal responses relative to the responses at position 0 deg, where the two test lines were superimposed in the RF center. In contrast, when the animal was performing the bisection task, the weak contextual inhibition was changed into strong facilitation. (Adapted from Crist, Li, & Gilbert, 2001.) (See color plate 5.)

training of monkeys in a three-line bisection discrimination task greatly enhances the animals' discrimination ability and markedly modifies contextual influences on V1 responses (Crist et al., 2001). The latter is characterized by an increase in the modulatory strength of contextual lines and, in some cases, a reversal of the modulatory effect from inhibition to facilitation (figure 8.3). That is, V1 neurons become more sensitive with training to positional displacement of parallel lines, an ability that is important in bisection discrimination. This change was present only in the trained retinotopic area while the monkey was doing the trained bisection task, suggesting that both encoding and retrieving the learned information require task-specific top-down influences.

While training on discrimination of simple stimuli can sharpen neuronal selectivity in early visual areas, learning to discriminate among complex shapes can enhance shape selectivity of neurons in the inferior temporal cortex (area IT), the last stage in cortical processing of visual objects (Logothetis, Pauls, & Poggio, 1995; Kobatake, Wang, & Tanaka, 1998; Freedman, Riesenhuber, Poggio, & Miller, 2006). This enhancement shows orientation dependency as well: neuronal selectivity was stronger for stimuli presented at the trained orientation than for rotated versions of the same stimuli (Logothetis et al., 1995; Freedman et al., 2006).

Enhanced neuronal responsiveness and shifted cortical representation A sharpening of neuronal selectivity seen in training on discrimination of delicate changes in the related stimulus attributes implies that fewer cells would respond to the stimulus, leading to an overall decrease in neuronal activity. This speculation has been supported by some imaging studies: training on orientation discrimination (Schiltz et al., 1999) and contrast discrimination (Mukai et al., 2007) reduces activation in visual cortical areas. This is in contrast with training on detection of low-saliency targets that are presented near the contrast detection threshold or are camouflaged within a noisy background. Neuronal responsiveness is usually enhanced specifically to the familiar target when the intensity or signal-to-noise ratio of the target is low.

In a study that trained monkeys to detect visual contours embedded in a complex background, striking parallel changes were observed in response properties of V1 neurons and the behavioral performance of the animals (W. Li, Piech, & Gilbert, 2008). As illustrated in figure 8.1, visual saliency of the embedded contour increases with the number of collinear lines forming the contour. For monkeys that have never been trained on the detection task, V1 neuronal responses are independent of the presence and length of the embedded contours, regardless of whether or not the animals' attention is directed to the target location (figure 8.4*A*). During the training, the animals' ability to detect the camouflaged contours increases gradually. Correspondingly, contour-related neuronal responses, which are closely correlated with the animals' performance on contour detection, are built up in V1 (figure 8.4*B*). Analyses based on signal detection theory indicate that responses of individual V1 neurons in well-trained animals are predictive of the animals' performance in the contour detection task (see also W. Li et al., 2006). Moreover, the learning effects are specific to the trained retinotopic location in terms of both the behavioral and neuronal responses. These findings highlight the importance of a specific combination of stimulus and task, or a specific interaction between the bottom-up and top-down processes, in inducing learning-associated cortical changes. In fact, the same set of interactions is important in the retrieving process, since doing a task irrelevant to contour detection significantly reduces neuronal responses to the embedded contours in the trained animals (figure 8.4*C*). Furthermore, a complete removal of any potential forms of top-down influences with anesthesia, which largely spares neuronal selectivity for basic stimulus attributes, completely abolishes the contour-related responses in V1 (figure 8.4*D*).

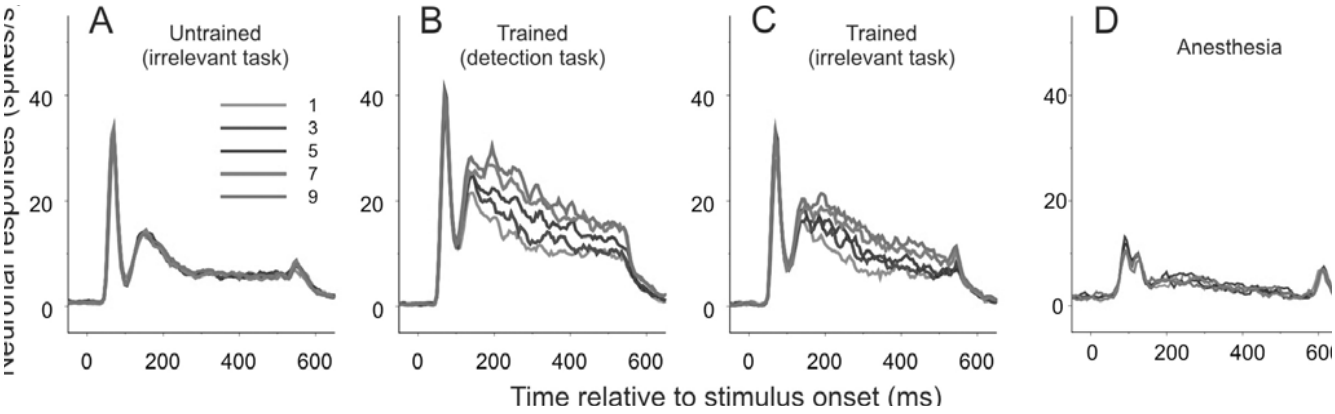


FIGURE 8.4 Learning- and task-dependent changes in V1 associated with training on contour detection. Shown here are averaged population neuronal responses to visual contours consisting of 1, 3, 5, 7, and 9 collinear lines embedded in an array of randomly oriented lines (for example see figure 8.1). Time 0 indicates stimulus onset. (*A*) Neuronal responses in V1 of untrained monkeys are independent of contour lengths (the six peristimulus time histograms are superimposed), indicating the absence of contour information in V1 responses. (*B*) Over the course of training the animals on contour detection, a late response component associated with contour saliency emerges—the longer the contours, the stronger the neuronal responses. (*C*) In trained animals the contour-related V1 responses are much weakened when the animals perform tasks that are irrelevant to contour detection. (*D*) Contour-related responses disappear in the trained V1 region under anesthesia. (Adapted from Li, Piech, & Gilbert, 2008.) (See color plate 6.)

Similar to contour integration, the detection of a difference in texture between a small area and a large area surrounding it involves the horizontal integrative mechanisms. A study showed that a single session of training on such a surface segmentation task increases fMRI signals in early visual areas (Schwartz, Maquet, & Frith, 2002). A further study has shown that the maximal increases occur in the first couple of weeks of training before the subjects' detection performance reaches a plateau (Yotsumoto, Watanabe, & Sasaki, 2008). With prolonged training, the elevated fMRI signals drop back to the levels before training. This result is opposite to the electrophysiological finding that for monkeys extensively trained on contour detection, the learning-induced neuronal responses in V1 are retained (W. Li et al., 2008).

Training on detection of an isolated target near contrast threshold can also selectively boost activity in early visual cortex. After training human subjects to detect a near-threshold grating patch, the fMRI signals in V1 are significantly increased for the trained orientation (Furmanski, Schluppeck, & Engel, 2004). Enhancement of neuronal responsiveness associated with detection training has also been demonstrated in higher cortical areas along the visual processing streams. For instance, training monkeys to identify natural scene images that are degraded by noise specifically enhances V4 neuronal responses to those familiar and degraded pictures (Rainer, Lee, & Logothetis, 2004). In detection of coherent motion of dynamic random dots, an improvement in monkeys' performance is correlated by enhanced neuronal responses in areas MT and MST (Zohary, Celebrini, Britten, & Newsome, 1994; but see Law & Gold, 2008, which argues that learning-associated improvement in detection of coherent motion does not involve changes in MT, but rather it largely relies on the stage that makes perceptual decisions).

Visual search can be taken as a special detection task in which a target is camouflaged in an array of similar distracters. Increased neuronal responsiveness in V1 has been reported to be associated with animals' familiarity with the target (Lee, Yang, Romero, & Mumford, 2002). In addition to heightened activity in early visual areas, learning to search for a simple geometric shape within distractors causes a concomitant decrease in fMRI signals in higher visual areas involved in shape processing (Sigman et al., 2005). This finding suggests that extensive training can shift cortical representation of the learned shape from higher to lower visual areas for more efficient and less effortful processing. This idea is further supported by the evidence that extensive training on a perceptual task significantly reduces activity in the frontoparietal cortical network for attentional control (Pollmann & Maertens, 2005; Sigman et al., 2005; Mukai et al., 2007).

Temporal code In addition to the firing rates, changes in temporal response properties of neurons have also been suggested to be related to perceptual learning. In the primary somatosensory cortex, neuronal responses become more coherent with training on tactile frequency discrimination. This change correlates better with the improved discrimination ability than does cortical recruitment (Recanzone, Merzenich, & Schreiner, 1992). Likewise, in

the primary auditory cortex, neuronal responses become more phase-locked to the trained and behavior-relevant acoustic pulses (Bao, Chang, Woods, & Merzenich, 2004). In the early visual areas of cats, gamma oscillations—an indication of response synchrony between neurons (Singer, 1999)—are increased for a behaviorally relevant visual stimulus on which the cats are trained (Salazar, Kayser, & Konig, 2004). An enhancement of coherence and synchrony in neuronal responses may reflect plastic changes at synaptic levels.

Task-specific top-down influences

Psychophysical studies reveal that perceptual learning usually does not happen simply by repeated passive exposure to a stimulus. For the same visual stimulus, a subject's performance on discrimination of a stimulus attribute can be improved only if the attribute is attended and used in the perceptual task (Shiu & Pashler, 1992; Ahissar & Hochstein, 1993; Saffell & Matthews, 2003). Moreover, the improvement does not generalize to discrimination of the other stimulus attributes of the same stimulus. Task specificity of perceptual learning indicates that top-down influences play an important role in encoding the learned information. Changes in neuronal response properties with training exhibit similar task dependency. As mentioned earlier (figure 8.4), when a naïve monkey is passively exposed to the embedded visual contours, no change in V1 responses is observed. Most importantly, the same is true when the animal attends to the target location but does a task irrelevant to contour detection, indicating that spatial attention by itself cannot differentiate neuronal responses to contours of different lengths. The contour-related responses emerge and get strengthened only when the animal starts learning the contour detection task. These findings suggest that top-down influences are not limited to spatial attention, but can convey much more information, including information about specific perceptual tasks.

Task-dependent modification of neuronal responses has been clearly demonstrated by a physiological study (W. Li, Piech, & Gilbert, 2004) in which monkeys were trained to perform either a bisection or a vernier discrimination task with an identical set of stimulus patterns (figure 8.5). Neuronal responses in V1 are strongly modulated by the stimulus attribute relevant to the immediate task, but they are little affected by the other task-irrelevant attribute.

Taken together, accumulated evidence indicates that V1 neurons take on novel response properties related to the perceptual task over the course of perceptual learning.

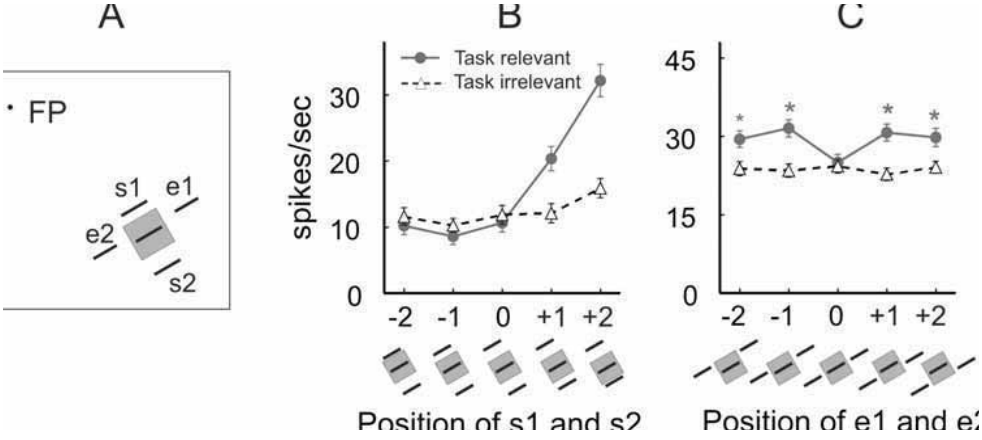


FIGURE 8.5 Task-specific top-down influences on V1 responses. (*A*) Monkeys were trained to do two different discrimination tasks with identical stimulus patterns at the same visual field location. The stimuli consisted of five simultaneously presented lines: an optimally oriented line fixed in the RF center and flanked by four additional lines surrounding the RF. In different trials, the arrangement of the two side flankers (s1, s2) was randomly assigned from a set of five different configurations (illustrated in the cartoons at the bottom of *B*, labeled from −2 to +2). Each configuration differs from the others in the separation between the three side-by-side lines (in condition 0 the three lines were equidistant; in the other conditions either s1 or s2 was closer to the central line). In the same trials, the two end-flankers (e1, e2) were also independently assigned a random configuration from a set of predefined arrangements, such that the end flankers were collinear with each other but misaligned with the central line to either side (the cartoons at the bottom of *C*). The animal was cued to perform either a bisection task based on the three side-by-side lines or a vernier task based on the three end-to-end lines, using the same set of five-line stimuli. (*B*) Responses of a V1 cell were examined as a function of the position of the two side flankers s1 and s2 when the animal either performed the bisection task, in which s1 and s2 were task-relevant; or performed the vernier task, in which the same s1 and s2 were task-irrelevant. (*C*) Responses of a V1 cell were examined as a function of the position of the two end flankers e1 and e2 when the animal either performed the vernier task, in which e1 and e2 were task-relevant; or performed the bisection task, in which the same e1 and e2 were task-irrelevant (Adapted from W. Li, Piech, & Gilbert, 2004.) (See color plate 7.)

Moreover, similar to so-called state-dependent learning (for example, see Shulz, Sosnik, Ego, Haidarliu, & Ahissar, 2000), retrieval of the acquired neuronal response properties requires a recurrence of the same stimulus and task used for the training. Task-dependent modification of neuronal response properties has also been reported in auditory cortex (for review see Fritz, Elhilali, & Shamma, 2005). This cortical mechanism can account for the stimulus and task specificity of perceptual learning. The information related to a given stimulus attribute is represented at the level of subsets of inputs to a cell, which are gated by the top-down signals via interactions between feedback connections from higher cortical areas and intrinsic connections within V1. This mechanism enables multiple attributes to be represented by the same cells without cross talk, greatly expanding the information-processing capability of neurons. The fast functional switching or multiplexing capability of visual neurons under task-specific top-down control is tightly coupled with perceptual learning, as repeated execution of the same perceptual task, and therefore, repetitive invoking of top-down influences specific to the task can potentiate the dynamic changes useful for solving the perceptual tasks, leading to encoding and retrieving of the implicit memory formed during perceptual learning.

Epilogue

Visual cortical plasticity is not limited to postnatal development and to contingent reactions induced by anomalous experiences. It is a lifelong ongoing process accompanying visual perception, as shown in various cortical changes associated with perceptual learning. There has been considerable debate about the neural basis of perceptual learning regarding the cortical loci where the plastic changes occur, since conflicting results are often reported. To derive an unbiased point of view based on the mixed results, one must take into account the nature of visual perception, which, according to Helmholtz, is nothing more than our subjective ideas or inference derived from sensory stimulation (Helmholtz, 1866). It is now evident that the generation of visual percepts depends on information processing distributed across a large number of cortical areas, such as the visual areas dedicated to sensory processing, the attentional network engaged in top-down control, and the executive network involved in making perceptual decisions. Therefore, it is not surprising that changes associated with perceptual learning could be observed in any of these cortical areas. Another complication comes from the variety of possible visual stimuli and tasks, as well as the limitation of individual approaches used in different studies. The classical fable of "the blind men and the elephant" is always good to keep in mind when considering the rigorous debate on perceptual learning and on cortical plasticity.

REFERENCES

Adini, Y., Wilkonsky, A., Haspel, R., Tsodyks, M., & Sagi, D. (2004). Perceptual learning in contrast discrimination: The effect of contrast uncertainty. *J. Vis.*, *4*, 993–1005.

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proc. Natl. Acad. Sci. USA*, *90*, 5718–5722.

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annu. Rev. Neurosci.*, *25*, 339–379.

Allman, J., Miezin, F., & McGuinness, E., (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu. Rev. Neurosci.*, *8*, 407–430.

Baker, C. I., Peli, E., Knouf, N., & Kanwisher, N. G. (2005). Reorganization of visual processing in macular degeneration. *J. Neurosci.*, *25*, 614–618.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, *218*, 697–698.

Bao, S., Chang, E. F., Woods, J., & Merzenich, M. M. (2004). Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. *Nat. Neurosci.*, *7*, 974–981.

Brown, M., Irvine, D. R. F., & Park, V. N. (2004). Perceptual learning on an auditory frequency discrimination task by cats: Association with changes in primary auditory cortex. *Cereb. Cortex*, *14*, 952–965.

Burkhalter, A., Bernardo, K. L., & Charles, V. (1993). Development of local circuits in human visual cortex. *J. Neurosci.*, *13*, 1916–1931.

Burton, H. (2003). Visual cortex activity in early and late blind people. *J. Neurosci.*, *23*, 4005–4011.

Calford, M. B., & Tweedale, R. (1988). Immediate and chronic changes in responses of somatosensory cortex in adult flying fox after digit amputation. *Nature*, *332*, 446–448.

Calford, M. B., Wang, C., Taglianetti, V., Waleszczyk, W. J., Burke, W., & Dreher, B. (2000). Plasticity in adult cat visual cortex (area 17) following circumscribed monocular lesions of all retinal layers. *J. Physiol.*, *524*(Pt. 2), 587–602.

Calford, M. B., Wright, L. L., Metha, A. B., & Taglianetti, V. (2003). Topographic plasticity in primary visual cortex is mediated by local corticocortical connections. *J. Neurosci.*, *23*, 6434–6442.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annu. Rev. Neurosci.*, *21*, 47–74.

Callaway, E. M., & Katz, L. C. (1990). Emergence and refinement of clustered horizontal connections in cat striate cortex. *J. Neurosci.*, *10*, 1134–1153.

Chino, Y. M., Kaas, J. H., Smith, E. L., Langston, A. L., & Cheng, H. (1992). Rapid reorganization of cortical maps in adult cats following restricted deafferentation in retina. *Vis. Res.*, *32*, 789–796.

Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *J. Neurophysiol.*, *78*, 2889–2894.

Crist, R. E., Li, W., & Gilbert, C. D. (2001). Learning to see: Experience and attention in primary visual cortex. *Nat. Neurosci.*, *4*, 519–525.

Darian-Smith, C., & Gilbert, C. D. (1995). Topographic reorganization in the striate cortex of the adult cat and monkey is cortically mediated. *J. Neurosci.*, *15*, 1631–1647.

Devor, M., & Wall, P. D. (1978). Reorganization of spinal cord sensory map after peripheral nerve injury. *Nature*, *276*, 75–76.

Devor, M., & Wall, P. D. (1981). Plasticity in the spinal cord sensory map following peripheral nerve injury in rats. *J. Neurosci.*, *1*, 679–684.

Dilks, D. D., Serences, J. T., Rosenau, B. J., Yantis, S., & McCloskey, M. (2007). Human adult cortical reorganization and consequent visual distortion. *J. Neurosci.*, *27*, 9585–9594.

Dosher, B. A., & Lu, Z. L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proc. Natl. Acad. Sci. USA*, *95*, 13988–13993.

Ellison, A., & Walsh, V. (1998). Perceptual learning in visual search: Some evidence of specificities. *Vis. Res.*, *38*, 333–345.

Eysel, U. T., Schweigart, G., Mittmann, T., Eyding, D., Qu, Y., Vandesande, F., et al. (1999). Reorganization in the visual cortex after retinal and cortical damage. *Restor. Neurol. Neurosci.*, *15*, 153–164.

Fawcett, S. L., Wang, Y.-Z., & Birch, E. E. (2005). The critical period for susceptibility of human stereopsis. *Invest. Ophthalmol. Vis. Sci.*, *46*, 521–525.

Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vis. Res.*, *23*, 145–150.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*, 43–44.

Fox, K., & Wong, R. L. (2005). A comparison of experience-dependent plasticity in the visual and somatosensory systems. *Neuron*, *48*, 465–477.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb. Cortex*, *16*, 1631–1644.

Fritz, J., Elhilali, M., & Shamma, S. (2005). Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear. Res.*, *206*, 159–176.

Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Curr. Biol.*, *14*, 573–578.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vis. Res.*, *41*, 711–724.

Ghose, G. M., Yang, T. M., & Maunsell, J. H. R. (2002). Physiological correlates of perceptual learning in monkey V1 and V2. *J. Neurophysiol.*, *87*, 1867–1888.

Giannikopoulos, D. V., & Eysel, U. T. (2006). Dynamics and specificity of cortical map reorganization after retinal lesions. *Proc. Natl. Acad. Sci. USA*, *103*, 10805–10810.

Gilbert, C. D. (1983). Microcircuitry of the visual cortex. *Annu. Rev. Neurosci.*, *6*, 217–247.

Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. [Review] [108 refs]. *Neuron*, *9*, 1–13.

Gilbert, C. D. (1998). Adult cortical dynamics. [Review.] [230 refs.] *Physiol. Rev.*, *78*, 467–485.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. [Review] [170 refs]. *Neuron*, *31*, 681–697.

Gilbert, C. D., & Wiesel, T. N. (1992). Receptive field dynamics in adult primary visual cortex. *Nature*, *356*, 150–152.

Han, Y. K., Kover, H., Insanally, M. N., Semerdjian, J. H., & Bao, S. (2007). Early experience impairs perceptual discrimination. *Nat. Neurosci.*, *10*, 1191–1197.

Hebb, D. O. (1949). *Organization of behavior: A neuropsychological theory*. New York: Wiley.

Heinen, S. J., & Skavenski, A. A. (1991). Recovery of visual responses in foveal V1 neurons following bilateral foveal lesions in adult monkey. *Exp. Brain Res.*, *83*, 670–674.

Helmholtz, H. (1866). *Treatise on physiological optics*, Vol. 3. New York: Dover Publications, 1962.

Hubel, D. H., & Wiesel, T. N. (1970). Period of susceptibility to physiological effects of unilateral eye closure in kittens. *J. Physiol. London*, *206*, 419–436.

Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.*, *198*, 1–59.

James, W. (1890). *The principles of psychology*. New York: Dover, 1950.

Jones, E. G. (1994). Santiago Ramón y Cajal and the Croonian Lecture, March 1894. *Trends Neurosci.*, *17*, 190–192.

Kaas, J. H., Krubitzer, L. A., Chino, Y. M., Langston, A. L., Polley, E. H., & Blair, N. (1990). Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina. *Science*, *248*, 229–231.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proc. Natl. Acad. Sci. USA*, *88*, 4966–4970.

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophysiol.*, *80*, 324–330.

Konorski, J. (1948). *Conditioned reflexes and neuron organization*. Cambridge, UK: Cambridge University Press.

Kovacs, I., Kozma, P., Feher, A., & Benedek, G. (1999). Late maturation of visual spatial integration in humans. *Proc. Natl. Acad. Sci. USA*, *96*, 12204–12209.

Law, C. T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.*, *11*, 505–513.

Lee, T. S., Yang, C. F., Romero, R. D., & Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nat. Neurosci.*, *5*, 589–597.

Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by retuning the decision "template" for position discrimination. *Nat. Neurosci.*, *7*, 178–183.

Li, W., & Gilbert, C. D. (2002). Global contour saliency and local colinear interactions. *J. Neurophysiol.*, *88*, 2846–2856.

Li, W., Piech, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.*, *7*, 651–657.

Li, W., Piech, V., & Gilbert, C. D. (2006). Contour saliency in primary visual cortex. *Neuron*, *50*, 951–962.

Li, W., Piech, V., & Gilbert, C. D. (2008). Learning to link visual contours. *Neuron*, *57*, 442–451.

Locke, J. (1689). *An essay concerning human understanding*. Amherst, NY: Prometheus Books, 1995.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.*, *5*, 552–563.

Masuda, Y., Dumoulin, S. O., Nakadomari, S., & Wandell, B. A. (2008). V1 projection zone signals in human macular degeneration depend on task, not stimulus. *Cereb. Cortex*, *18*, 2483–2493.

McKee, S. P., & Westheimer, G. (1978). Improvement in vernier acuity with practice. *Percept. Psychophys.*, *24*, 258–262.

Merzenich, M. M., Kaas, J. H., Wall, J., Nelson, R. J., Sur, M., & Felleman, D. (1983a). Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience*, *8*, 33–55.

Merzenich, M. M., Kaas, J. H., Wall, J. T., Sur, M., Nelson, R. J., & Felleman, D. J. (1983b). Progression of change following median nerve section in the cortical representation of the hand in areas 3b and 1 in adult owl and squirrel monkeys. *Neuroscience*, *10*, 639–665.

Merzenich, M. M., Nelson, R. J., Stryker, M. P., Cynader, M. S., Schoppmann, A., & Zook, J. M. (1984). Somatosensory cortical map changes following digit amputation in adult monkeys. *J. Comp. Neurol.*, *224*, 591–605.

Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., & Ungerleider, L. G. (2007). Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *J. Neurosci.*, *27*, 11401–11411.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annu. Rev. Neurosci.*, *28*, 377–401.

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, *256*, 1018–1021.

Pollmann, S., & Maertens, M. (2005). Shift of activity from attention to motor-related brain areas during visual learning. *Nat. Neurosci.*, *8*, 1494–1496.

Pons, T. P., Garraghty, P. E., Ommaya, A. K., Kaas, J. H., Taub, E., & Mishkin, M. (1991). Massive cortical reorganization after sensory deafferentation in adult macaques. *Science*, *252*, 1857–1860.

Raiguel, S., Vogels, R., Mysore, S. G., & Orban, G. A. (2006). Learning to see the difference specifically alters the most informative V4 neurons. *J. Neurosci.*, *26*, 6589–6602.

Rainer, G., Lee, H., & Logothetis, N. K. (2004). The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biol.*, *2*, E44.

Rajan, R., Irvine, D. R. F., Wise, L. Z., & Heil, P. (1993). Effect of unilateral partial cochlear lesions in adult cats on the representation of lesioned and unlesioned cochleas in primary auditory cortex. *J. Comp. Neurol.*, *338*, 17–49.

Ramachandran, V. S., & Braddick, O. (1973). Orientation-specific learning in stereopsis. *Perception*, *2*, 371–376.

Ramón y Cajal, S. (1911). *Histologie du système nerveux de l'homme et des vertébrés*. Madrid: Consejo Superior de Investigaciones Cientificas, reprinted 1972.

Rasmusson, D. D. (1982). Reorganization of raccoon somatosensory cortex following removal of the 5th digit. *J. Comp. Neurol.*, *205*, 313–326.

Recanzone, G. H., Merzenich, M. M., & Jenkins, W. M. (1992). Frequency discrimination training engaging a restricted skin surface results in an emergence of a cutaneous response zone in cortical area 3A. *J. Neurophysiol.*, *67*, 1057–1070.

Recanzone, G. H., Merzenich, M. M., Jenkins, W. M., Grajski, K. A., & Dinse, H. R. (1992). Topographic reorganization of the hand representation in cortical area 3B of owl monkeys trained in a frequency-discrimination task. *J. Neurophysiol.*, *67*, 1031–1056.

Recanzone, G. H., Merzenich, M. M., & Schreiner, C. E. (1992). Changes in the distributed temporal response properties of S1 cortical neurons reflect improvements in performance on a temporally based tactile discrimination task. *J. Neurophysiol.*, *67*, 1071–1091.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.*, *13*, 87–103.

Robertson, D., & Irvine, D. R. F. (1989). Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness. *J. Comp. Neurol.*, *282*, 456–471.

Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: What is learned? *Vis. Res.*, *35*, 519–527.

Saffell, T., & Matthews, N. (2003). Task-specific perceptual learning on speed and direction discrimination. *Vis. Res.*, *43*, 1365–1374.

Salazar, R. F., Kayser, C., & Konig, P. (2004). Effects of training on neuronal activity and interactions in primary and higher visual cortices in the alert cat. *J. Neurosci.*, *24*, 1627–1636.

Schiltz, C., Bodart, J. M., Dubois, S., Dejardin, S., Michel, C., Roucoux, A., et al. (1999). Neuronal mechanisms of perceptual learning: Changes in human brain activity with training in orientation discrimination. *Neuroimage*, *9*, 46–62.

Schmid, L. M., Rosa, M. G. P., Calford, M. B., & Ambler, J. S. (1996). Visuotopic reorganization in the primary visual cortex of adult cats following monocular and binocular retinal lesions. *Cereb. Cortex*, *6*, 388–405.

Schoups, A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularity. *J. Physiol. (Lond.)*, *483*, 797–810.

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, *412*, 549–553.

Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: A functional MRI study of visual texture discrimination. *Proc. Natl. Acad. Sci. USA.*, *99*, 17137–17142.

Series, P., Latham, P. E., & Pouget, A. (2004). Tuning curve sharpening for orientation selectivity: Coding efficiency and the impact of correlations. *Nat. Neurosci.*, *7*, 1129–1135.

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Percept. Psychophys.*, *52*, 582–588.

Shulz, D. E., Sosnik, R., Ego, V., Haidarliu, S., & Ahissar, E. (2000). A neuronal analogue of state-dependent learning. *Nature*, *403*, 549–553.

Sigman, M., Cecchi, G. A., Gilbert, C. D., & Magnasco, M. O. (2001). On a common circle: Natural scenes and Gestalt rules. *Proc. Natl. Acad. Sci. USA*, *98*, 1935–1940.

Sigman, M., & Gilbert, C. D. (2000). Learning to find a shape. *Nat. Neurosci.*, *3*, 264–269.

Sigman, M., Pan, H., Yang, Y. H., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron*, *46*, 823–835.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*, 49–65.

Sireteanu, R., & Rettenbach, R. (2000). Perceptual learning in visual search generalizes over tasks, locations, and eyes. *Vis. Res.*, *40*, 2925–2949.

Sireteanu, R., & Rieth, C. (1992). Texture segregation in infants and children. *Behav. Brain Res.*, *49*, 133–139.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.*, *27*, 279–306.

Stettler, D. D., Das, A., Bennett, J., & Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, *36*, 739–750.

Sugita, Y. (1996). Global plasticity in adult visual cortex following reversal of visual input. *Nature*, *380*, 523–526.

Teich, A. F., & Qian, N. (2003). Learning and adaptation in a recurrent model of V1 orientation selectivity. *J. Neurophysiol.*, *89*, 2086–2100.

Vaina, L. M., Belliveau, J. W., Roziers, E. B., & Zeffiro, T. A. (1998). Neural systems underlying learning and representation of global motion. *Proc. Natl. Acad. Sci. USA*, *95*, 12657–12662.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, *255*, 419–423.

Vogels, R., & Orban, G. A. (1985). The effect of practice on the oblique effect in line orientation judgments. *Vis. Res.*, *25*, 1679–1687.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Percept. Psychophys.*, *56*, 495–500.

Weiss, T., Miltner, W. H. R., Liepert, J., Meissner, W., & Taub, E. (2004). Rapid functional plasticity in the primary somatomotor cortex and perceptual changes after nerve block. *Eur. J. Neurosci.*, *20*, 3413–3423.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychol. Forsch.*, *4*, 301–350.

Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol.*, *26*, 1003–1017.

Yang, T., & Maunsell, J. H. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *J. Neurosci.*, *24*, 1617–1626.

Yotsumoto, Y., Watanabe, T., & Sasaki, Y. (2008). Different dynamics of performance and brain activation in the time course of perceptual learning. *Neuron*, *57*, 827–833.

Yu, C., Klein, S. A., & Levi, D. M. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *J. Vis.*, *4*, 169–182.

Zohary, E., Celebrini, S., Britten, K. H., & Newsome, W. T. (1994). Neuronal plasticity that underlies improvement in perceptual performance. *Science*, *263*, 1289–1292.

9 Characterizing and Modulating Neuroplasticity of the Adult Human Brain

ALVARO PASCUAL-LEONE

ABSTRACT Neurons are highly specialized structures, are resistant to change, but are engaged in distributed networks that do dynamically change over the lifespan. Changes in functional connectivity, for example by shifts in synaptic strength, can be followed by more stable structural changes. Therefore, the brain is continuously undergoing plastic remodeling. Plasticity is not an occasional state of the nervous system but is the normal ongoing state of the nervous system throughout the lifespan. It is not possible to understand normal psychological function or the manifestations or consequences of disease without invoking the concept of brain plasticity. The challenge is to understand the mechanisms and consequences of plasticity in order to modulate them, suppressing some and enhancing others, in order to promote adaptive brain changes. Behavioral, neurostimulation, and targeted neuropharmacological interventions can modulate plasticity and promote desirable outcomes for a given individual.

Human behavior is molded by environmental changes and pressures, physiological modifications, and experiences. The brain, as the source of human behavior, must thus have the capacity to dynamically change in response to shifting afferent inputs and efferent demands. However, individual neurons are highly complex and exquisitely optimized cellular elements, and their capacity for change and modification is necessarily very limited. Fortunately, these stable cellular elements are engaged into neural networks that assure functional stability while providing a substrate for rapid adaptation to shifting demands. Dynamically changing neural networks might thus be considered evolution's invention to enable the nervous system to escape the restrictions of its own genome (and its highly specialized cellular specification) and adapt fluidly and promptly to environmental pressures, physiological changes, and experiences.

Therefore, representation of function in the brain may be best conceptualized by the notion of distributed neural networks, a series of assemblies of neurons (nodes) that might be widely dispersed anatomically but are structurally interconnected, and that can be functionally integrated to serve a specific behavioral role. Such nodes can be conceptualized as operators that contribute a given computation independent of the input ("metamodal brain"; see Pascual-Leone & Hamilton, 2001). However, the computations at each node might also be defined by the inputs themselves. Inputs shift depending on the integration of a node in a distributed neural network, and the layered and reticular structure of the cortex with rich reafferent loops provides the substrate for rapid modulation of the engaged network nodes. Depending on behavioral demands, neuronal assemblies can be integrated into *different* functional networks by shifts in weighting of connections (functional and effective connectivity). Indeed, *timing* of interactions between elements of a network, beyond integrity of structural connections, might be a critical binding principle for the functional establishment of given network action and behavioral output. Such notions of dedicated, but multifocal, networks, which can dynamically shift depending on demands for a given behavioral output, provide a current resolution to the long-standing dispute between localizationists and equipotential theorists. Function comes to be identified with a certain pattern of activation of specific, spatially distributed, but interconnected neuronal assemblies in a specific time window and temporal order. In such distributed networks, specific nodes may be critical for a given behavioral outcome. Knowledge of such instances is clinically useful to explain findings in patients and localize their lesions, but it provides an oversimplified conceptualization of brain-behavior relations. In the setting of dynamically plastic neural networks, behavior following an insult is never simply the result of the lesion, but rather the consequence of how the rest of the brain is capable of sustaining function following a given lesion. Neural plasticity can confer no perceptible change in the behavioral output of the brain, cannot lead to changes demonstrated only under special testing conditions, and cannot cause behavioral changes that constitute symptoms of disease. There may be loss of a previously acquired

ALVARO PASCUAL-LEONE Berenson-Allen Center for Noninvasive Brain Stimulation, Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts

behavioral capacity, release of behaviors normally suppressed in the uninjured brain, takeover of lost function by neighboring systems (albeit perhaps incompletely or by means of different strategies and computations), or emergence of new behaviors that may prove adaptive or maladaptive for the individual.

Therefore, plasticity is not an occasional state of the nervous system; instead, it is the normal ongoing state of the nervous system throughout the life span. A full, coherent account of any motor, sensory, or cognitive theory has to build into its framework the fact that the nervous system undergoes continuous changes in response to modifications in its input afferents and output targets. It is not possible to understand normal psychological function or the manifestations or consequences of disease without invoking the concept of brain plasticity. However, plasticity at the neural level does not speak to the question of behavioral change and certainly does not necessarily imply functional recovery or even functional change. The challenge we face is to learn enough about the mechanisms of plasticity and the mapping relations between brain activity and behavior to be able to guide them, suppressing changes that may lead to undesirable behaviors while accelerating or enhancing those that result in a behavioral benefit for the subject or patient.

Activity across neural networks changes dynamically to preserve behavior

As long as an output pathway to manifest the behavior is preserved (even if alternate pathways need to be unmasked or facilitated), changes in the activity across a distributed neural network may be able to establish new patterns of brain activation and sustain function, even in the face of a focal insult.

This dynamic changing capacity of neural networks is illustrated by the following experiment (Pascual-Leone, Amedi, Fregni, & Merabet, 2005). Normal subjects were asked to open and close their fist deliberately at a self-paced rhythm of approximately one movement every second while lying in an fMRI scanner. As compared with rest, during movement there was a significant activation of the primary motor cortex (M1) contralateral to the moving hand and of the rostral supplementary motor cortex (SMA) (figure 9.1*A*). If motor cortex activity is modified by repetitive transcranial magnetic stimulation (TMS), the pattern of brain activation changes while behavioral integrity is maintained (figure 9.1*A*). Slow, repetitive TMS (rTMS) suppresses activity in the targeted cortical area (Valero-Cabre, Payne, & Pascual-Leone, 2007; Valero-Cabre, Payne, Rushmore, Lomber, & Pascual-Leone, 2005). After slow rTMS has been applied to the contralateral M1, the subjects are able to sustain behavior, but they reveal an increased activation of the rostral SMA and of M1 ipsilateral to the moving hand. Conversely, increasing excitability in the contralateral M1 (by application of fast rTMS; Valero-Cabre et al., 2007, 2005) is associated with a decrease in activation of rostral SMA.

Lee and colleagues (2003), combining TMS and positron emission tomography (PET), have revealed the shifts in cortico-cortical and cortico-subcortical connectivity underlying the changes in cortical activation patterns that are associated with sustained behavior in the face of focal cortical disruption (figure 9.1*B*). Following repetitive TMS (rTMS), motor performance remained unchanged while task-dependent increases in regional cerebral blood flow (rCBF) were seen during movement in the directly stimulated M1 and the dorsal premotor cortex. Analyses of effective connectivity showed that after rTMS there is a remodeling of the motor system, with increased movement-related connectivity from the SMA and premotor cortex to sites in primary sensorimotor cortex.

Thus, in the face of an imposed disruption of focal brain activity, performance of a relatively simple task can be maintained by rapid operational remapping of representations, recruitment of additional brain areas, and task-related changes in cortico-cortical and cortico-muscular coherence (Chen et al., 2003; Lee et al., 2003; Oliviero, Strens, Di Lazzaro, Tonali, & Brown, 2003; Strens, Fogelson, Shanahan, Rothwell, & Brown, 2003). However, under other circumstances, modulation of activity in a focal node of a distributed neural network can give rise to changes in behavior in a controlled and specific manner. For example, a right parietal lesion can result in spatial neglect (the failure to explore contralesional space), yet remarkably, the neglect symptoms can completely and abruptly disappear following a second lesion to the left frontal cortex (Vuilleumier, Hester, Assal, & Regli, 1996) (figure 9.2*A*). Such a paradoxical effect of a brain lesion (Kapur, 1996), resulting in a behavioral improvement, is consistent with animal studies by Sprague (Sprague, 1966) and later Payne and Lomber (Payne, Lomber, Geeraerts, van der Gucht, & Vandenbussche, 1996). Given reciprocal interhemispheric inhibition and the proposed link to attentional performance, suppression of one parietal cortex may lead to contralateral neglect, but at the same time, the disinhibition of structures involved in interhemispheric competition might lead to a functional release in the opposite hemisphere, which could result in a measurable ipsilateral behavioral enhancement. Indeed, Hilgetag, Theoret, and Pascual-Leone (2001) provided experimental support for such notions. Normal subjects had to detect small rectangular stimuli briefly presented on a computer monitor either unilaterally in the left or right periphery, or bilaterally in both. Spatial detection performance was tested before and immediately after a 10-minute, 1-Hz rTMS train to (a) right parietal cortex; (b) left parietal

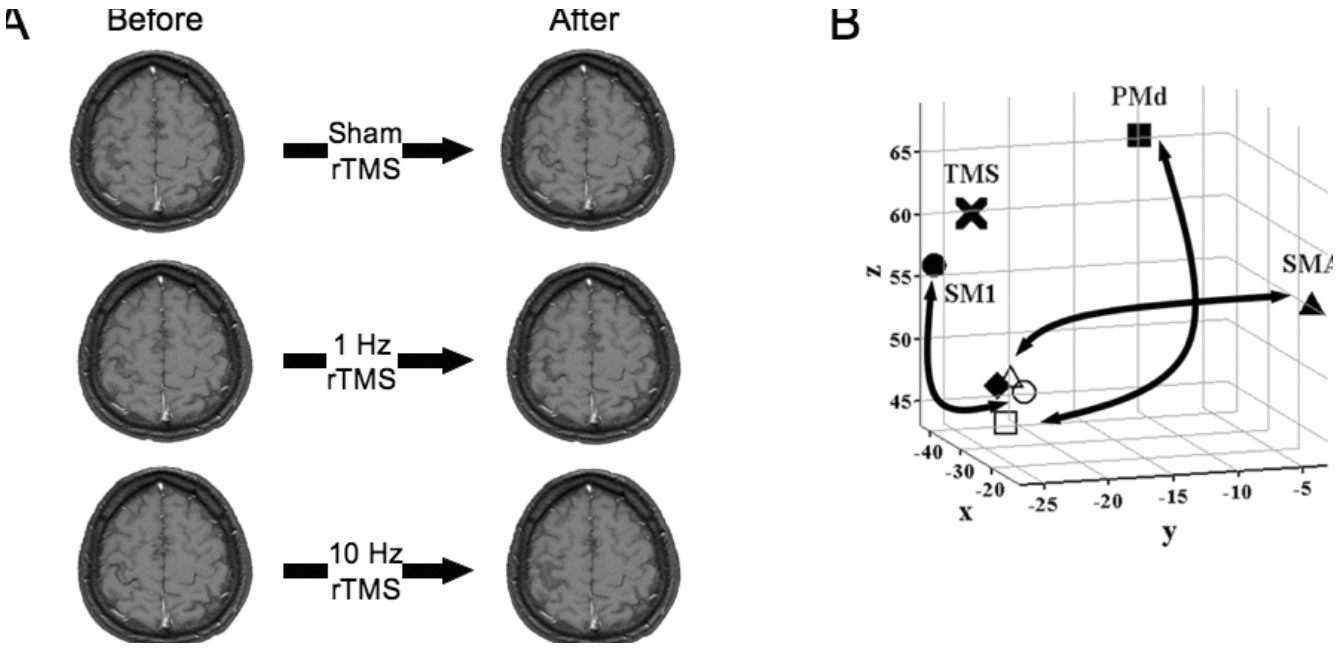


FIGURE 9.1 (*A*) Brain activation in fMRI while subjects performed the same rhythmic hand movement (under careful kinematic control) before and after repetitive transcranial magnetic stimulation (rTMS) of the contralateral motor cortex. Following sham rTMS (*top row*) there is no change in the significant activation of the motor cortex (M1) contralateral to the moving hand and of the rostral supplementary motor cortex (SMA). After M1 activity is suppressed using 1-Hz rTMS (1,600 stimuli, 90% of motor threshold intensity; *middle row*), there is an increased activation of the rostral SMA and of M1 ipsilateral to the moving hand. Increasing excitability in the contralateral M1 using high-frequency rTMS (20 Hz, 90% of motor threshold intensity, 1,600 stimuli; *bottom row*) results in a decrease in activation of rostral SMA. (See color plate 8.) (*B*) Areas of the brain showing differential movement-related responses and coupling after rTMS. Circle, square, and triangle symbols indicate sites in primary motor cortex (*open symbols*) that are more strongly coupled to activity in sensorimotor cortex (SM1), dorsal premotor cortex (PMd), and supplementary motor cortex (SMA) during a finger movement task after rTMS. *X* marks the site of stimulation with 1-Hz rTMS. (*B*) modified from Lee and colleagues (2003).

cortex; (c) right primary motor cortex; and (d) sham stimulation. Hilgetag and colleagues observed a clear extinction phenomenon for stimuli presented contralaterally to the stimulated hemisphere (right or left parietal cortex). However, the deficit was accompanied by increased detection for unilateral stimuli presented on the side of the stimulated hemisphere compared to baseline (figure 9.2*B*). None of the control stimulation sites had any effect on the detection performance. These insights can be translated to parietal-damaged patients with neglect, in whom rTMS to the undamaged (frequently left) hemisphere can alleviate hemi-inattention symptoms (Brighina et al., 2003; Oliveri et al., 1999).

Therefore, activity in neural networks is dynamically modulated, and this fact can be illustrated by the neurophysiological adaptations to focal brain disruptions or lesions. Behavioral outcome, however, does not map in a fixed manner to changes in activity in distributed networks. Thus changes in network activity can give rise to no behavioral change, behavioral improvements, or losses. The frequently held notion that the brain optimizes behavior is therefore not correct, for it implies, for example, that a lesion to the brain will always lead to a loss rather than enhancement of function. In fact, we have seen that this view is challenged by the conceptualization of the brain as endowed with dynamic plasticity.

However, the scope of possible dynamic changes across a given neural network is defined by existing connections. Genetically controlled aspects of brain development define neuronal elements and initial patterns of connectivity. Given such initial, genetically determined, individually different brain substrates, the same events will result in diverse consequences as plastic brain mechanisms act upon individually distinct neural substrates. Similarly, within each individual, differences across neural networks (e.g., visual system, auditory system, or language system) will also condition the range of plastic modification (Bavelier & Neville, 2002; Neville & Bavelier, 2002). Plastic changes across brain systems vary as a function of differences in patterns of existing connections and in molecular and genetically controlled factors across brain systems that define the range, magnitude, stability, and chronometry of plasticity.

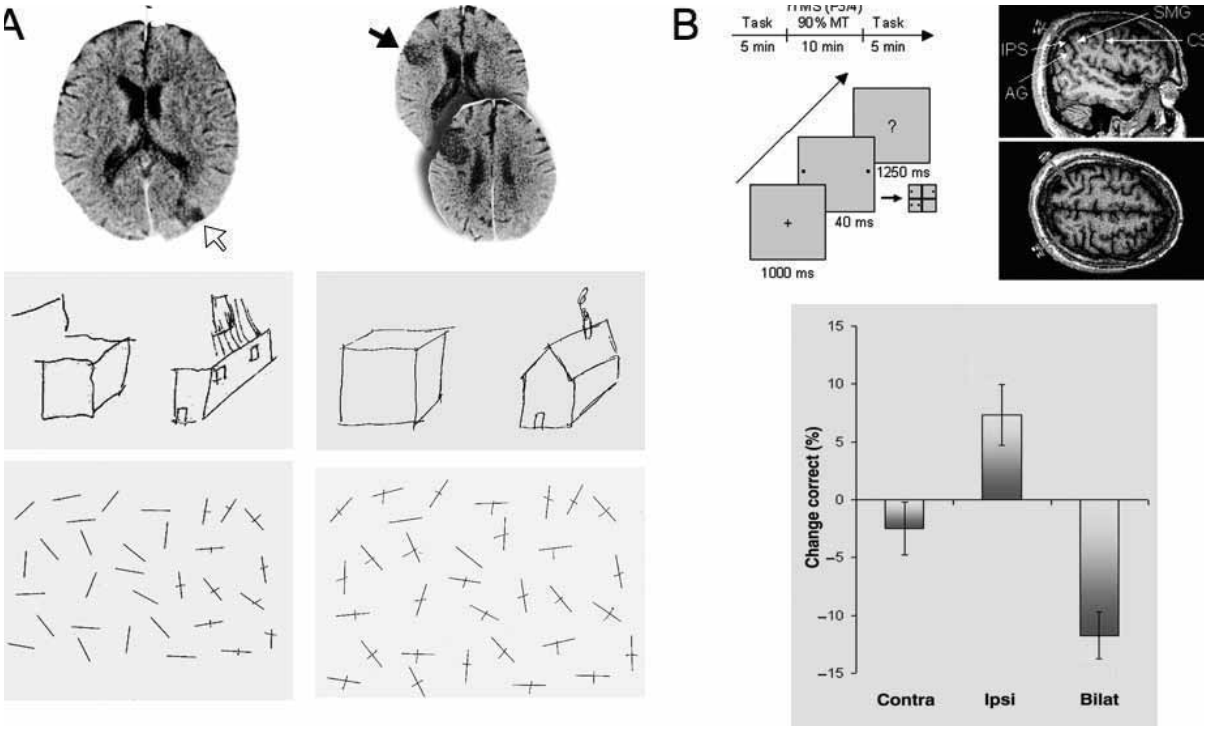


FIGURE 9.2 (*A*) Findings of neglect in house-drawing and line-cancellation tasks (*left panel*) due to a right parietal stroke (*open arrow*). The signs of neglect acutely resolved (*right panel*) as a consequence of a second stroke in the left, frontal cortex (*filled arrow*). (Modified from Vuilleumier, Hester, Assal, & Regli, [1996].) (*B*) Impact of image-guided rTMS to the right parietal cortex on a visual stimulus detection task. During the task subjects were presented with carefully titrated visual stimuli on the right, left, or bilateral side of a computer monitor (*top left*) and had to respond by pressing the appropriate response button (right, left, or both). TMS was applied guided by the subject's own anatomical MRI using a frameless stereotaxic system (*top right*). There was a decrease in contralateral performance (neglect) but an even greater increase in performance ipsilateral to the parietal rTMS location (*bottom*). This summed up to a significant decrease in bilateral stimuli, where subjects neglected the contralateral stimulus and responded as if only the ipsilateral one had been presented (extinction of double simultaneous stimulation). (*B* modified from Hilgetag, Theoret, & Pascual-Leone, [2001]). (See color plate 9.)

Dynamic network changes can be followed by more stable plastic changes

Rapid, ongoing changes in neural networks in response to environmental influences (for example, by dynamic shifts in the strength of preexisting connections across distributed neural networks, changes in task-related cortico-cortical and cortico-subcortical coherence, or modifications of the mapping between behavior and neural activity) may be followed by the establishment of new connections through dendritic growth and arborization resulting in structural changes and establishment of new pathways.

These two steps of plasticity are illustrated by the following experiment (Pascual-Leone et al., 1995). Normal subjects were taught to perform with one hand a five-finger exercise on a piano keyboard connected to a computer through a musical interface. They were instructed to perform the sequence of finger movements fluently, without pauses, and without skipping any keys, while paying particular attention to keeping the interval between the individual key presses constant and the duration of each key press the same. A metronome gave a tempo of 60 beats per minute for which the subjects were asked to aim, as they performed the exercise under auditory feedback. Subjects were studied on five consecutive days, and each day they had a two-hour practice session followed by a test. The test consisted of the execution of 20 repetitions of the five-finger exercise. The number of sequence errors decreased, and the duration, accuracy, and variability of the intervals between key pushes (as marked by the metronome beats) improved significantly over the course of the five days. Before the first practice session on the first day of the experiment and daily thereafter, we used TMS to map the motor cortical areas targeting long finger flexor and extensor muscles bilaterally. As the subjects' performance improved, the threshold for TMS activation of the finger flexor and extensor muscles decreased steadily. Even considering this change in threshold, the size of the cortical representation for both muscle groups increased significantly (figure 9.3*A*, Week 1). Remarkably, this increase in size of the cortical output maps could be demonstrated only when the cortical mapping studies were conducted shortly after the practice session, but no longer the next day, after a night

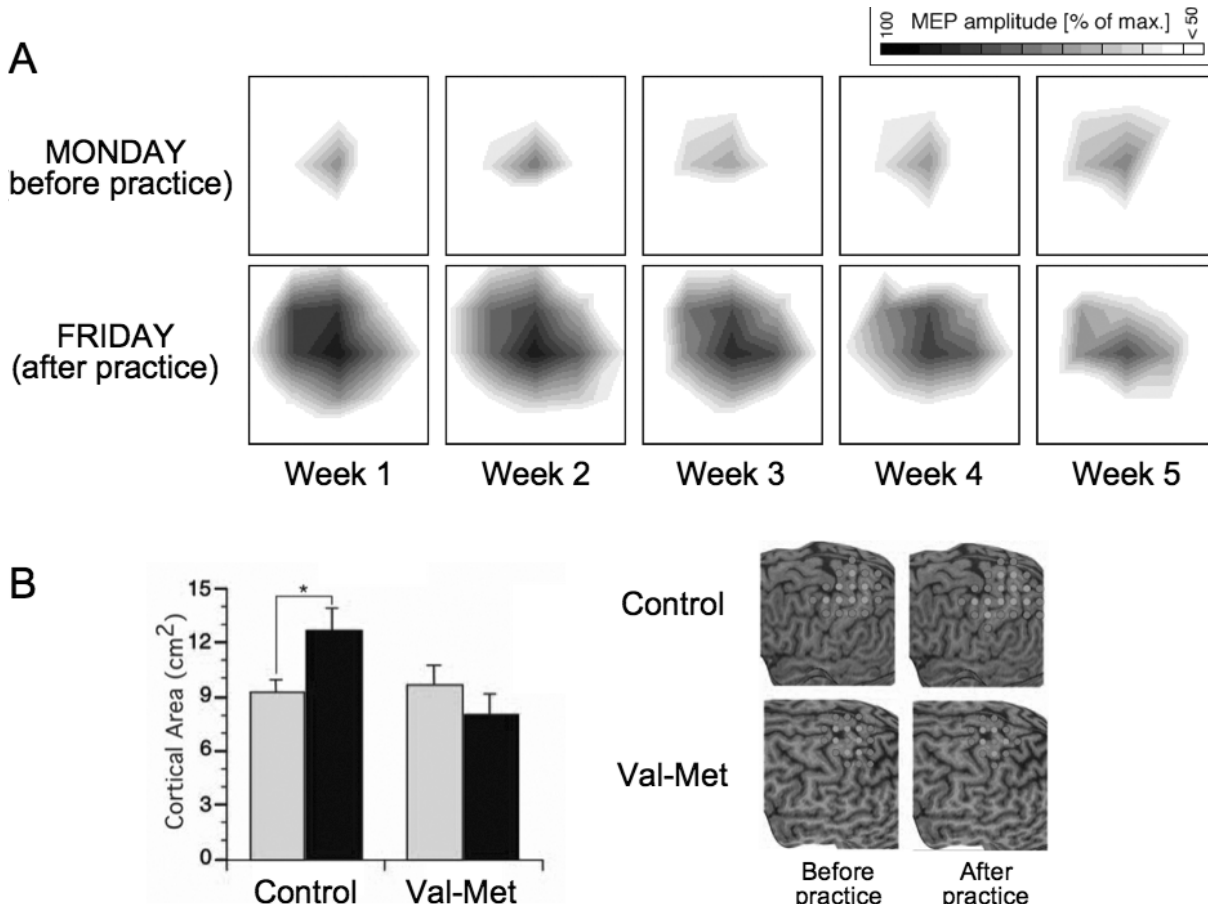


FIGURE 9.3 (*A*) Cortical output maps for the finger flexors during acquisition of a five-finger movement exercise on a piano. There are marked changes of the output maps for finger flexors of the trained hand over the five weeks of daily practice (Monday to Friday). Note that there are two distinct processes in action, one accounting for the rapid modulation of the maps from Mondays to Fridays and the other responsible for the slow and more discrete changes in Monday maps over time. (Modified from Pascual-Leone, 1996; Pascual-Leone et al., 1995.) (*B*) Histogram displaying the size of the cortical output maps before (gray bars) and after exercise (black bars) in control subjects and subjects with a val-66met polymorphism for BDNF (*left side*). Following exercise, control subjects had significantly larger representations than at baseline, whereas subjects with a Met allele did not show a significant change. This difference is further illustrated by the representative motor maps from control and Val-Met polymorphism subjects superimposed onto a composite brain MRI image of the cortex (*right side*). Sites from which TMS evoked criterium responses in the target muscle are marked in green; negative sites are marked in red. (Modified from Kleim et al., 2006.) (See color plate 10.)

of sleep and before the next day's practice session. Interestingly, even such initial steps of experience- and practice-related plasticity seem critically regulated by genetic factors. Kleim and colleagues (2006) used TMS to map cortical motor output and show that training-dependent changes in motor-evoked potentials and motor map organization are reduced in subjects with a val66met polymorphism in the brain-derived neurotrophic factor (BDNF) gene, as compared to subjects without the polymorphism (figure 9.3*B*).

Once a near-perfect level of performance was reached at the end of a week of daily practice, subjects continued daily practice of the same piano exercise during the following four weeks (Group 1) or stopped practicing (Group 2) (Pascual-Leone, 1996). During the four weeks of follow-up (figure 9.3*A*, Weeks 2–5), cortical output maps for finger flexor and extensor muscles were obtained in all subjects on Mondays (before the first practice session of that week in Group 1) and on Fridays (after the last practice session for the week in Group 1). In the group that continued practicing (Group 1), the cortical output maps obtained on Fridays showed an initial peak and eventually a slow decrease in size despite continued performance improvement. However, the maps obtained on Mondays, before the practice session and following the weekend rest, showed a small change from baseline with a tendency to increase in size over the course of the study. In Group 2, the maps returned to baseline after the first week of follow-up and remained stable thereafter.

This experiment illustrates two distinct phases of modulation of motor output maps. The rapid time course in the initial modulation of the motor outputs, by which a certain region of motor cortex can reversibly increase its influence on a motoneuron pool, is most compatible with

the unmasking of previously existing connections. Supporting this notion, the initial changes are quite transient: demonstrable after practice, but returning to baseline after a weekend rest. As the task becomes overlearned over the course of five weeks, the pattern of cortical activation for optimal task performance changes as other neural structures take a more leading role in task performance. Flexible, short-term modulation of existing pathways represents a first and necessary step leading up to longer-term structural changes in the intracortical and subcortical networks as skills become overlearned and automatic. A growing number of neuroimaging studies have suggested a similar two-step process (Grafton et al., 1992; Jenkins, Brooks, Nixon, Frackowiak, & Passingham, 1994; Karni et al., 1995, 1998; Seitz, Roland, Bohm, Greitz, & Stone, 1990), and animal studies support the notion of different processes involved, over time, in early acquisition and later consolidation of skill learning (Kleim et al., 2004).

Two complementary mechanisms control plasticity

As indicated in the preceding section, dynamic network changes can lead to more stable plastic changes, which involve synaptic plasticity as well as dendritic arborization and network remodeling. Such changes might be conceptualized as the result of a balance between two complimentary mechanisms—one promoting and the other limiting plasticity (figure 9.4). Both these mechanisms are critical in assuring that appropriate synapses are formed and unnecessary synapses are pruned in order to optimize functional systems necessary for cognition and behavior. Though the molecular mechanisms that contribute to plasticity are numerous and complex, the plasticity-promoting mechanism appears to be critically dependent on the neurotrophin BDNF (brain-derived neurotrophic factor) (Lu, 2003), while genes within the major histocompatibility complex (MHC) Class I appear to be involved in the plasticity-limiting mechanism (Boulanger, Huh, & Shatz, 2001; Huh et al., 2000).

At the synaptic level, mechanisms of long-term potentiation (LTP) and long-term depression (LTD) involve a series of induction and consolidation steps that are dependent on various structural changes and can be modified, increased, or suppressed by distinct modulatory influences (Lynch, Rex, & Gall, 2007). LTP is initiated by the influx of calcium through glutamate receptors in the postsynaptic density. Calcium-activated kinases and proteinases disassemble the cytoskeleton, made up of actin filaments cross-linked by spectrin and other proteins, that normally maintains the shape of the dendritic spines. Thus the spine becomes rounder and shorter, effectively enlarging the surface of the postsynaptic density, which can then accept a greater number of glutamate receptors and provide better access to proteins

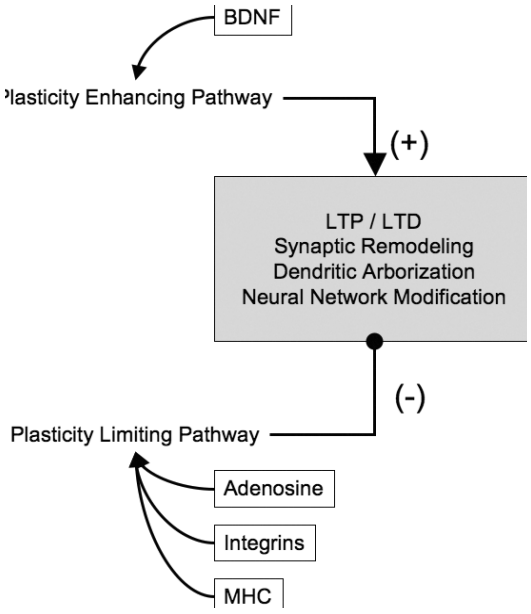


FIGURE 9.4 A schematic diagram of the conceptualization of plasticity as the balance of plasticity-enhancing and plasticity-limiting mechanisms, which are dependent on different neuromodulators.

that enhance current flow through the receptors. In parallel, signaling from adhesion receptors, particularly integrins, and modulatory receptors, particularly BDNF, induces the rapid polymerization of actin and the formation of a new cytoskeleton. This polymerization of actin filaments consolidates the new dendritic spine morphology and thus the LTP. Despite the complexity of such a process and the numerous molecules involved, BDNF appears to be the most potent enhancer of plasticity discovered thus far, playing a critical role in LTP consolidation across multiple brain regions. BDNF has been shown to facilitate LTP in the visual cortex (Akaneya, Tsumoto, Kinoshita, & Hatanaka, 1997) and the hippocampus (Korte et al., 1995). At CA1 synapses, a weak tetanic stimulation, which in and of itself would only induce short-term potentiation of low magnitude, leads to strong LTP when paired with BDNF (Figurov, Pozzo-Miller, Olafsson, Wang, & Lu, 1996). During motor training, BDNF levels are elevated within motor cortex (Klintsova, Dickson, Yoshida, & Greenough, 2004), and human subjects who have a single nucleotide polymorphism in the BDNF gene (val66met) show reduced experience-dependent plasticity of the motor cortex following a voluntary motor task (Kleim et al., 2006).

In contrast, adenosine (Arai, Kessler, & Lynch, 1990) and ligands for integrins (Staubli, Vanderklish, & Lynch, 1990) block LTP when applied immediately after theta burst stimulation because of the disruption of actin polymerization and LTP consolidation. Along these lines, a blind screen for

genes involved in normal developmental activity-dependent remodeling of neuronal connectivity revealed a region of DNA better known for its role in immune functioning, namely Class I major histocompatibility complex (Class I MHC) (Corriveau, Huh, & Shatz, 1998). More recent studies suggest that MHC Class I genes are an integral part of an experience-dependent plasticity-limiting pathway (Syken, Grandpre, Kanold, & Shatz, 2006). Such negative modulators of synaptic plasticity are needed. Establishing and strengthening new synapses is an important part of developmental plasticity, but this has to be coupled with normal regressive events including activity-dependent synaptic weakening and elimination of inappropriate connections. Without these regressive events, superfluous synapses may persist and may impair normal neural development.

Therefore, different modulators, including BDNF on the one side and adenosine or MHC Class I genes on the other, serve complimentary functions that lead to the development and rapid modulation of functional circuits across the whole brain. Such dynamic systems do harbor potential dangers, and disruption of these pathways or their relative balance may lead to severe pathological states. However, these opposing pathways offer the opportunity for interventions and thus for guiding plasticity for the benefit of individual subjects.

Plasticity as the cause of disease

Focal hand dystonia (Quartarone, Siebner, & Rothwell, 2006) may be a good example of pathological consequences of plasticity that can be promoted by suitable genetic predispositions, such as DYT-1 or others. Importantly, though, the mere induction of certain plastic changes is not sufficient to lead to disability. Similar plastic changes can be documented in patients with focal dystonia and proficient musicians (Quartarone et al., 2006; Rosenkranz, Williamon, & Rothwell, 2007). Furthermore, musicians can develop focal hand dystonia (Chamagne, 2003), and the underlying pathophysiology appears to be slightly different than in other forms of dystonia, such as writer's cramp (Rosenkranz et al., 2008). Perhaps "faulty" practice or excessive demand in the presence of certain predisposing factors may result in unwanted cortical rearrangement and lead to disease. It seems clear, though, that plastic changes in the brain do not speak to behavioral impact. Similar changes can be associated with behavioral advantages (as in the professional musicians) or neurological disability (as in the case of focal dystonia), presumably on the basis of modulatory influences from distributed neural activity.

Chronic, neuropathic pain syndromes have also been argued to represent "pathological" consequences of plasticity (Flor, 2008; Fregni, Pascual-Leone, & Freedman, 2007; Zhuo, 2008). Tinnitus may be the result of plasticity in the auditory system induced by abnormal cochlear input (Bartels, Staal, & Albers, 2007). Schizophrenia, depression, posttraumatic stress disorder, and attention-deficit/hyperactivity disorder are all conditions that may, in part, represent disorders of brain plasticity (Frost et al., 2004; Hayley, Poulter, Merali, & Anisman, 2005; Rapoport & Gogtay, 2008). Drug addiction and perhaps addictive behaviors in general are argued to represent examples of pathology as the consequence of plasticity (Kalivas & O'Brien, 2008; Kauer & Malenka, 2007). Alzheimer's disease appears to be linked to abnormal synaptic plasticity that may in fact constitute a crucial initial step in the pathogenesis of the disease (Selkoe, 2008). Autism may be another example of plasticity-mediated pathology: genetic factors may lead to a predisposition such that developmentally mediated plasticity (possibly in itself controlled by abnormal regulators) results in pathological complex behaviors affecting social interactions, language acquisition, or sensory processing (Morrow et al., 2008).

Therefore, human behavior and the manifestations of human disease are ultimately heavily defined by brain plasticity. An initial, genetically determined neural substrate is modified during development and environmental interactions by plasticity. The processes of neural plasticity themselves can be normal, but may act upon an abnormal nervous system as a consequence of genetic or specific environmental factors. Alternatively, the mechanisms of plasticity themselves may be abnormal, potentially compounding the consequences of an abnormal substrate on the basis of a genetically determined "starting point" or environmental insult. In any case, interventions to guide behavior or treat pathological symptomatology might be more immediate in their behavioral repercussions and thus more effective if aimed at modulating plasticity than if intent on addressing underlying genetic predispositions.

Fragile X syndrome provides a suitable illustration for such notions (Bear, Dolen, Osterweil, & Nagarajan, 2008; O'Donnell & Warren, 2002; Penagarikano, Mulle, & Warren, 2007). The genetic mutation responsible for fragile X syndrome, *FMR1*, leads to the absence of the encoded protein FMRP, which appears to play an important role in synaptic plasticity by regulating metabotropic-glutamate-receptor-dependent LTD. Thus in the absence of FMRP there is excessive experience-dependent LTD. Mouse models of fragile X syndrome have also demonstrated impairments in LTP, possibly as a result of immature development of dendritic spines. However, the application of BDNF to slices from FMR1 knockout mice fully restores LTP to normal levels (Lauterborn et al., 2007), and thus it might be possible to normalize cognitive function and behavior in patients with fragile X by pharmacologically "normalizing" the affected mechanisms of plasticity.

Plasticity as an opportunity for intervention

The plastic nature of the brain provides, following injury, a risk of maladaptive change and perpetuation of deficits, but also an opportunity for intervention and overcoming of symptoms. Following brain injury, behavior (regardless of whether normal or manifesting injury-related deficits) remains the consequence of the functioning of the entire brain, and thus the consequence of a plastic nervous system. Ultimately, symptoms are not the manifestation of the injured brain region, but rather the expression of plastic changes in the rest of the brain. Following an insult, cortico-cortical and cortico-subcortico-cortical interactions will shift weights across the involved neural network, aiming to adapt to the functional disruption and establish a suitable brain activation map for a desired behavioral result. Different mechanisms, which may proceed partly in parallel but which have variable time frames, are likely involved. Initial plastic changes aim to minimize damage. Dysfunctional, but not damaged, neuronal elements may recover from the postinjury shock and penumbra processes. Partially damaged neural elements may be able to be repaired relatively quickly after the insult as well, thus contributing to early functional improvement. Subsequent processes, once the final damage has been established, involve relearning (rather than recovery) and may, as we have discussed, follow a two-step process: initial unmasking and strengthening of existing neural pathways, and eventually the establishment of new structural changes. At all these stages of plastic adaptation, neurostimulation and targeted neuropharmacological interventions may be able to guide the neural processes and promote adaptive, desirable outcomes for a given individual.

Consider, for example, the recovery of hand motor function following a stroke (Alonso-Alonso, Fregni, & Pascual-Leone, 2007; Cramer & Riley, 2008; Di Filippo et al., 2008; Nudo, 2006). After stroke, there is an increase in the excitability of the unaffected hemisphere, presumably owing to reduced transcallosal inhibition from the damaged hemisphere and increased use of the intact hemisphere. For example, in patients with acute cortical stroke, intracortical inhibition is decreased and intracortical facilitation increased in the unaffected hemisphere (Liepert, Storch, Fritsch, & Weiller, 2000). Furthermore, the interhemispheric inhibitory drive from the unaffected to the affected motor cortex in the process of voluntary movement generation is abnormal (Murase, Duque, Mazzocchio, & Cohen, 2004), and this imbalance of excitability between the hemispheres is inversely correlated with the time since the stroke (Shimizu et al., 2002).

Acutely after a stroke, increased inhibitory input from the undamaged to the damaged hemisphere makes conceptual sense if one considers it a manifestation of a neural attempt to control perilesional activity, reduce oxygen and glucose demands in the penumbra of the stroke, and thus limit the extension of the lesion. However, after an acute phase, and once the injury is stable, input to the perilesional area would seem to be best as excitatory in nature to maximize the capability of the preserved neurons in the injured tissue to drive behavioral output. If so, following the acute phase, we might expect a shift of interhemispheric (and many intrahemispheric) interactions from inhibitory to excitatory. Should such a shift fail to take place, the resulting functional outcome may be undesirable, with limited behavioral restoration, in part owing to persistent inhibitory inputs from the intact to the damaged hemisphere. In fact, some neuroimaging studies demonstrate that long-term, persistent activation of the ipsilateral cortex during motor tasks is associated with poor motor outcomes, whereas a good motor recovery is associated with a decrease in activity in the unaffected area and an increase in the affected primary sensorimotor cortex activity (Fregni & Pascual-Leone, 2006; Rossini et al., 2007; Ward & Cohen, 2004). If so, neuromodulatory approaches targeting the intact hemisphere may be useful to limit injury and promote recovery after a stroke. For instance, suppression of the ipsilateral motor cortex through slow rTMS may enhance motor performance in patients stable following the acute phase of a stroke (figure 9.5). In patients 1–2 months after a stroke, Mansur and colleagues (2005) applied 0.5 Hz rTMS for 10 min to the unaffected hemisphere to suppress cortical activity and thus release the damaged hemisphere from potentially excessive transcallosal inhibition. The results of this study support the notion that the overactivity of the unaffected hemisphere (ipsilateral hemisphere) may hinder hand-function recovery, and neuromodulation can be an interventional tool to accelerate this recovery. However, Werhahn, Conforto, Kadom, Hallett, and Cohen (2003) conducted a similar study to evaluate the modulation effects of 1 Hz rTMS of the unaffected hemisphere on the paretic hand and found different results. In that study, 1 Hz rTMS of the unaffected hemisphere did not affect the finger tapping in the paretic hand in a small sample of five patients more than one year after a stroke. The time since the brain insult is likely to be a critical variable to consider.

Of course, the alternative neuromodulatory approach, directly aimed at enhancing excitability of the damaged hemisphere perilesionally, can also be entertained. Khedr, Ahmed, Fathy, and Rothwell (2005) have reported extremely encouraging results along these lines. Similar principles of neuromodulation can be applied to the recovery of nonmotor strokes and other focal brain lesions as illustrated by studies on the effects of cortical stimulation on neglect discussed earlier (Brighina et al., 2003; Hilgetag et al., 2001; Oliveri et al., 1999) or the experience with aphasia (Martin

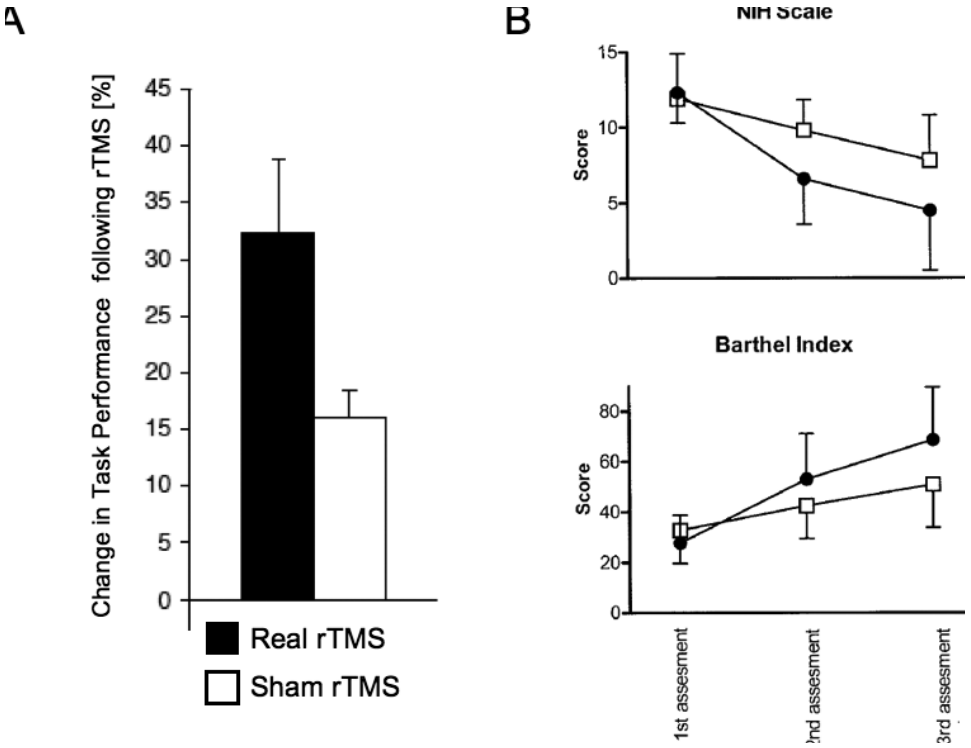


FIGURE 9.5 (*A*) Histogram illustrates the significant improvement in performance of the Purdue Pegboard task in stroke patients (on average 12 months after the stroke) following real (but not sham) slow-frequency repetitive transcranial magnetic stimulation (rTMS) to the unaffected hemisphere to decrease interhemispheric inhibition of the lesioned hemisphere and improve motor function. (Modified from Mansur et al., 2005.) (*B*) Serial assessments in patients with acute ischemic strokes undergoing 10 days of daily sessions of real or sham, fast rTMS over the affected motor cortex. Disability scales (Barthel Index and NIH Stroke Scale) measured before rTMS, at the end of the last rTMS session, and 10 days later show that real rTMS (filled symbols) improved patients' scores significantly more than sham (open symbols). (Modified from Khedr, Ahmed, Fathy, & Rothwell, 2005.)

et al., 2004; Naeser et al., 2005). However, challenges for such approaches remain, as our understanding of the various issues involved and how to optimize and individualize the neuromodulatory interventions is still rather sketchy. In any case, neuromodulatory approaches based on brain stimulation techniques are certainly not the only potential avenues to guide plasticity with therapeutic intent. Behavioral interventions, including technology-supported approaches, such as robotic or computerized task training, as well as pharmacological methods, might be equally effective.

A most intriguing question to consider is the possibility of similarly modulating plasticity in the attempt to promote functional gains in normal subjects (Canli et al., 2007; de Jongh, Bolt, Schermer, & Olivier, 2008; Farah et al., 2004; Lanni et al., 2008). Might it, for example, be possible to promote skill acquisition or verbal or nonverbal learning by enhancing certain plastic processes and suppressing others? This type of question raises important ethical issues, but also offers the potential for interventions that might be applicable in educational settings and translationally to patients. For example, consistent with the findings in recovery of hand motor function after a stroke, noninvasive cortical stimulation that suppresses excitability in the M1 ipsilateral or enhances excitability in the M1 contralateral to a training hand might result in varying degrees of improvement in motor function in healthy humans. Anodal transcranial direct current stimulation (tDCS) applied over M1 to increase its excitability before or during practice can lead to improvements in implicit motor learning as measured with the serial reaction time task (Nitsche et al., 2003), performance of a visuomotor coordination task (Antal et al., 2004) and a sequential finger movement task (Vines, Nair, & Schlaug, 2006), and performance of the Jebsen Taylor Hand function test (JTT) (Boggio et al., 2006). Similarly, the application of 1-Hz rTMS to suppress excitability of M1 ipsilateral to a training hand results in improvements in motor sequence learning (Kobayashi, Hutchinson, Théoret, Schlaug, & Pascual-Leone, 2004). However such effects might be task and condition specific. For example, learning of a more complex finger tracking task was not modified by the same 1 Hz rTMS to suppress excitability of M1 ipsilateral to a training hand (Carey, Fregni, & Pascual-Leone, 2006), and the beneficial effects of anodal tDCS to the contralateral hand in the JTT were limited to the nondominant hand in young healthy adults and the elderly (Boggio et al., 2006).

Conclusions

The brain is highly plastic, and that plasticity represents evolution's invention to enable the nervous system to escape the restrictions of its own genome (and its highly specialized cellular specification) and adapt to rapidly shifting and often unpredictable environmental and experiential changes. Plastic changes may not necessarily represent a behavioral gain for a given subject. Instead, plasticity may be as much a cause of pathology and disease as it is the substrate for skill acquisition, learning, environmental adaptation, and recovery from insult. Recovery of function after a focal brain injury, such as a stroke, is essentially learning with a partially disrupted neural network and illustrates the dangers and opportunities of such a plastic brain. We might conceive of a two-step process of plasticity, with initial rapid modulation of connectivity across neural networks possibly followed by more stable, structural changes. Both these steps can be conceptualized as regulated by distinct plasticity-enhancing and plasticity-suppressing mechanisms that may account for different pathologies, but also offer targets for neuromodulatory and targeted neuropharmacological interventions to promote shifts in brain-behavior mapping that might be most adaptive for a given individual.

ACKNOWLEDGMENTS Work on this study was supported by the Harvard-Thorndike General Clinical Research Center at BIDMC (NCRR MO1 RR01032); NIH grants K24 RR018875, R01-MH069898, RO1-DC05672, RO1-NS 47754, RO1-NS 20068, RO1-EY12091, R01-EB 005047; and the Nancy Lurie Marks Family Foundation.

REFERENCES

Akaneya, Y., Tsumoto, T., Kinoshita, S., & Hatanaka, H. (1997). Brain-derived neurotrophic factor enhances long-term potentiation in rat visual cortex. *J. Neurosci.*, *17*, 6707–6716.

Alonso-Alonso, M., Fregni, F., & Pascual-Leone, A. (2007). Brain stimulation in poststroke rehabilitation. *Cerebrovasc. Dis.*, *24* (Suppl. 1), 157–166.

Antal, A., Nitsche, M. A., Kincses, T. Z., Kruse, W., Hoffmann, K. P., & Paulus, W. (2004). Facilitation of visuomotor learning by transcranial direct current stimulation of the motor and extrastriate visual areas in humans. *Eur. J. Neurosci.*, *19*(10), 2888–2892.

Arai, A., Kessler, M., & Lynch, G. (1990). The effects of adenosine on the development of long-term potentiation. *Neurosci. Lett.*, *119*, 41–44.

Bartels, H., Staal, M. J., & Albers, F. W. (2007). Tinnitus and neural plasticity of the brain. *Otol. Neurotol.*, *28*(2), 178–184.

Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: Where and how? *Nat. Rev. Neurosci.*, *3*(6), 443–452.

Bear, M. F., Dolen, G., Osterweil, E., & Nagarajan, N. (2008). Fragile X: Translation in action. *Neuropsychopharmacology*, *33*(1), 84–87.

Boggio, P. S., Castro, L. O., Savagim, E. A., Braite, R., Cruz, V. C., Rocha, R. R., et al. (2006). Enhancement of non-dominant hand motor function by anodal transcranial direct current stimulation. *Neurosci. Lett.*, *404*(1–2), 232–236.

Boulanger, L. M., Huh, G. S., & Shatz, C. J. (2001). Neuronal plasticity and cellular immunity: Shared molecular mechanisms. *Curr. Opin. Neurobiol.*, *11*, 568–578.

Brighina, F., Bisiach, E., Oliveri, M., Piazza, A., La Bua, V., et al. (2003). 1 Hz repetitive transcranial magnetic stimulation of the unaffected hemisphere ameliorates contralesional visuospatial neglect in humans. *Neurosci. Lett.*, *336*, 131–133.

Canli, T., Brandon, S., Casebeer, W., Crowley, P. J., Du Rousseau, D., Greely, H. T., et al. (2007). Neuroethics and national security. *Am. J. Bioethics*, *7*(5), 3–13.

Carey, J. R., Fregni, F., & Pascual-Leone, A. (2006). rTMS combined with motor learning training in healthy subjects. *Restor. Neurol. Neurosci.*, *24*(3), 191–199.

Chamagne, P. (2003). Functional dystonia in musicians: Rehabilitation. *Hand Clin.*, *19*, 309–316.

Chen, W. H., Mima, T., Siebner, H. R., Oga, T., Hara, H., et al. (2003). Low-frequency rTMS over lateral premotor cortex induces lasting changes in regional activation and functional coupling of cortical motor areas. *Clin. Neurophysiol.*, *114*, 1628–1637.

Corriveau, R. A., Huh, G. S., & Shatz, C. J. (1998). Regulation of Class I MHC gene expression in the developing and mature CNS by neural activity. *Neuron*, *21*, 505–520.

Cramer, S. C., & Riley, J. D. (2008). Neuroplasticity and brain repair after stroke. *Curr. Opin. Neurol.*, *21*(1), 76–82.

de Jongh, R., Bolt, I., Schermer, M., & Olivier, B. (2008). Botox for the brain: Enhancement of cognition, mood and pro-social behavior and blunting of unwanted memories. *Neurosci. Biobehav. Rev.*, *32*(4), 760–776.

Di Filippo, M., Tozzi, A., Costa, C., Belcastro, V., Tantucci, M., Picconi, B., et al. (2008). Plasticity and repair in the post-ischemic brain. *Neuropharmacology*, *55*(3), 353–362.

Farah, M. J., Illes, J., Cook-Deegan, R., Gardner, H., Kandel, E., King, P., et al. (2004). Neurocognitive enhancement: What can we do and what should we do? *Nat. Rev. Neurosci.*, *5*(5), 421–425.

Figurov, A., Pozzo-Miller, L. D., Olafsson, P., Wang, T., & Lu, B. (1996). Regulation of synaptic responses to high-frequency stimulation and LTP by neurotrophins in the hippocampus. *Nature*, *381*, 706–709.

Flor, H. (2008). Maladaptive plasticity, memory for pain and phantom limb pain: Review and suggestions for new therapies. *Expert Rev. Neurother.*, *8*(5), 809–818.

Fregni, F., & Pascual-Leone, A. (2006). Hand motor recovery after stroke: Tuning the orchestra to improve hand motor function. *Cogn. Behav. Neurol.*, *19*(1), 21–33.

Fregni, F., Pascual-Leone, A., & Freedman, S. D. (2007). Pain in chronic pancreatitis: A salutogenic mechanism or a maladaptive brain response? *Pancreatology*, *7*(5–6), 411–422.

Frost, D. O., Tamminga, C. A., Medoff, D. R., Caviness, V., Innocenti, G., & Carpenter, W. T. (2004). Neuroplasticity and schizophrenia. *Biol Psychiatry*, *56*(8), 540–543.

Grafton, S. T., Mazziota, J. C., Presty, S., Friston, K. J., Frackowiak, R. S. J., & Phleps, M. E. (1992). Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *J. Neurosci.*, *12*, 2542–2548.

Hayley, S., Poulter, M. O., Merali, Z., & Anisman, H. (2005). The pathogenesis of clinical depression: Stressor- and cytokine-induced alterations of neuroplasticity. *Neuroscience*, *135*(3), 659–678.

Hilgetag, C. C., Theoret, H., & Pascual-Leone, A. (2001). Enhanced visual spatial attention ipsilateral to rTMS-induced "virtual lesions" of human parietal cortex. *Nat. Neurosci.*, *4*(9), 953–957.

Huh, G. S., Boulanger, L. M., Du, H., Riquelme, P. A., Brotz, T. M., & Shatz, C. J. (2000). Functional requirement for Class I MHC in CNS development and plasticity. *Science*, *290*, 2155–2159.

Jenkins, I. H., Brooks, D. J., Nixon, P. D., Frackowiak, R. S., & Passingham, R. E. (1994). Motor sequence learning: A study with positron emission tomography. *J. Neurosci.*, *14*, 3775–3790.

Kalivas, P. W., & O'Brien, C. (2008). Drug addiction as a pathology of staged neuroplasticity. *Neuropsychopharmacology*, *33*(1), 166–180.

Kapur, N. (1996). Paradoxical functional facilitation in brain-behaviour research: A critical review. *Brain*, *119*(Pt. 5), 1775–1790.

Karni, A., Meyer, G., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature*, *377*, 155–158.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., et al. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proc. Natl. Acad. Sci. USA*, *95*, 861–868.

Kauer, J. A., & Malenka, R. C. (2007). Synaptic plasticity and addiction. *Nat. Rev. Neurosci.*, *8*(11), 844–858.

Khedr, E. M., Ahmed, M. A., Fathy, N., & Rothwell, J. C. (2005). Therapeutic trial of repetitive transcranial magnetic stimulation after acute ischemic stroke. *Neurology*, *65*(3), 466–468.

Kleim, J. A., Chan, S., Pringle, E., Schallert, K., Procaccio, V., Jimenez, R., et al. (2006). BDNF val66met polymorphism is associated with modified experience-dependent plasticity in human motor cortex. *Nat. Neurosci.*, *9*(6), 735–737.

Kleim, J. A., Hogg, T. M., VandenBerg, P. M., Cooper, N. R., Bruneau, R., & Remple, M. (2004). Cortical synaptogenesis and motor map reorganization occur during late, but not early, phase of motor skill learning. *J. Neurosci.*, *24*(3), 628–633.

Klintsova, A. Y., Dickson, E., Yoshida, R., & Greenough, W. T. (2004). Altered expression of BDNF and its high-affinity receptor TrkB in response to complex motor learning and moderate exercise. *Brain Res.*, *1028*, 92–104.

Kobayashi, M., Hutchinson, S., Théoret, H., Schlaug, G., & Pascual-Leone, A. (2004). Repetitive TMS of the motor cortex improves ipsilateral sequential simple finger movements. *Neurology*, *62*(1), 91–98.

Korte, M., Carroll, P., Wolf, E., Brem, G., Thoenen, H., & Bonhoeffer, T. (1995). Hippocampal long-term potentiation is impaired in mice lacking brain-derived neurotrophic factor. *Proc. Natl. Acad. Sci. USA*, *12*, 8856–8860.

Lanni, C., Lenzken, S. C., Pascale, A., Del Vecchio, I., Racchi, M., Pistoia, F., et al. (2008). Cognition enhancers between treating and doping the mind. *Pharmacol. Res.*, *57*(3), 196–213.

Lauterborn, J. C., et al. (2007). Brain-derived neurotrophic factor rescues synaptic plasticity in a mouse model of fragile X syndrome. *J. Neurosci.*, *27*, 10685–10694.

Lee, L., Siebner, H. R., Rowe, J. B., Rizzo, V., Rothwell, J. C., Frackowiak, R. S., et al. (2003). Acute remapping within the motor system induced by low-frequency repetitive transcranial magnetic stimulation. *J. Neurosci.*, *23*(12), 5308–5318.

Liepert, J., Storch, P., Fritsch, A., & Weiller, C. (2000). Motor cortex disinhibition in acute stroke. *Clin. Neurophysiol.*, *111*, 671–676.

Lu, B. (2003). BDNF and activity-dependent synaptic modulation. *Learn. Memory*, *10*, 86–98.

Lynch, G., Rex, C. S., & Gall, C. M. (2007). LTP consolidation: Substrates, explanatory power, and functional significance. *Neuropharmacology*, *52*, 12–23.

Mansur, C. G., Fregni, F., Boggio, P. S., Riberto, M., Gallucci-Neto, J., Santos, C. M., et al. (2005). A sham stimulation-controlled trial of rTMS of the unaffected hemisphere in stroke patients. *Neurology*, *64*(10), 1802–1804.

Martin, P. I., Naeser, M. A., Theoret, H., Tormos, J. M., Nicholas, M., Kurland, J., et al. (2004). Transcranial magnetic stimulation as a complementary treatment for aphasia. *Sem. Speech Lang.*, *25*(2), 181–191.

Morrow, E. M., Yoo, S. Y., Flavell, S. W., Kim, T. K., Lin, Y., Hill, R. S., et al. (2008). Identifying autism loci and genes by tracing recent shared ancestry. *Science*, *321*(5886), 218–223.

Murase, N., Duque, J., Mazzocchio, R., & Cohen, L. G. (2004). Influence of interhemispheric interactions on motor function in chronic stroke. *Ann. Neurol.*, *55*(3), 400–409.

Naeser, M. A., Martin, P. I., Nicholas, M., Baker, E. H., Seekins, H., Kobayashi, M., et al. (2005). Improved picture naming in chronic aphasia after TMS to part of right Broca's area: An open-protocol study. *Brain Lang.*, *93*(1), 95–105.

Neville, H., & Bavelier, D. (2002). Human brain plasticity: Evidence from sensory deprivation and altered language experience. *Prog. Brain Res.*, *138*, 177–188.

Nitsche, M. A., Schauenburg, A., Lang, N., Liebetanz, D., Exner, C., Paulus, W., et al. (2003). Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. *J. Cogn. Neurosci.*, *15*(4), 619–626.

Nudo, R. J. (2006). Mechanisms for recovery of motor function following cortical damage. *Curr. Opin. Neurobiol.*, *16*(6), 638–644.

O'Donnell, W. T., & Warren, S. T. (2002). A decade of molecular studies of fragile X syndrome. *Annu. Rev. Neurosci.*, *25*, 315–338.

Oliveri, M., Rossini, P. M., Traversa, R., Cicinelli, P., Filippi, M. M., Pasqualetti, P., et al. (1999). Left frontal transcranial magnetic stimulation reduces contralesional extinction in patients with unilateral right brain damage. *Brain*, *122*, 1731–1739.

Oliviero, A., Strens, L. H., Di Lazzaro, V., Tonali, P. A., & Brown, P. (2003). Persistent effects of high frequency repetitive TMS on the coupling between motor areas in the human. *Exp. Brain Res.*, *149*, 107–113.

Pascual-Leone, A. (1996). Reorganization of cortical motor outputs in the acquisition of new motor skills. In J. Kinura & H. Shibasaki (Eds.), *Recent advances in clinical neurophysiology* (pp. 304–308). Amsterdam: Elsevier Science.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annu. Rev. Neurosci.*, *28*, 377–401.

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Prog. Brain. Res.*, *134*, 427–445.

Pascual-Leone, A., Nguyet, D., Cohen, L. G., Brasil-Neto, J. P., Cammarota, A., & Hallett, M. (1995). Modulation of muscle responses evoked by transcranial magnetic stimulation during the acquisition of new fine motor skills. *J. Neurophysiol*, *74*(3), 1037–1045.

Payne, B. R., Lomber, S. G., Geeraerts, S., van der Gucht, E., & Vandenbussche, E. (1996). Reversible visual hemineglect. *Proc. Natl. Acad. Sci. USA*, *93*, 290–294.

Penagarikano, O., Mulle, J. G., & Warren, S. T. (2007). The pathophysiology of fragile X syndrome. *Annu. Rev. Genomics Hum. Genet.*, *8*, 109–129.

Quartarone, A., Siebner, H. R., & Rothwell, J. C. (2006). Task-specific hand dystonia: Can too much plasticity be bad for you? *Trends Neurosci.*, *29*(4), 192–199.

Rapoport, J. L., & Gogtay, N. (2008). Brain neuroplasticity in healthy, hyperactive and psychotic children: Insights from neuroimaging. *Neuropsychopharmacology*, *33*(1), 181–197.

Rosenkranz, K., Butler, K., Williamon, A., Cordivari, C., Lees, A. J., & Rothwell, J. C. (2008). Sensorimotor reorganization by proprioceptive training in musician's dystonia and writer's cramp. *Neurology*, *70*(4), 304–315.

Rosenkranz, K., Williamon, A., & Rothwell, J. C. (2007). Motorcortical excitability and synaptic plasticity is enhanced in professional musicians. *J. Neurosci.*, *27*(19), 5200–5206.

Rossini, P. M., Altamura, C., Ferreri, F., Melgari, J. M., Tecchio, F., Tombini, M., et al. (2007). Neuroimaging experimental studies on brain plasticity in recovery from stroke. *Europa Medicophysica*, *43*(2), 241–254.

Seitz, R. J., Roland, E., Bohm, C., Greitz, T., & Stone, E. S. (1990). Motor learning in man: A positron emission tomographic study. *Neuroreport*, *1*, 57–60.

Selkoe, D. J. (2008). Soluble oligomers of the amyloid beta-protein impair synaptic plasticity and behavior. *Behav. Brain Res.*, *192*(1), 106–113.

Shimizu, T., Hosaki, A., Hino, T., Sato, M., Komori, T., Hirai, S., et al. (2002). Motor cortical disinhibition in the unaffected hemisphere after unilateral cortical stroke. *Brain*, *125* (Pt. 8), 1896–1907.

Sprague, J. M. (1966). Interaction of cortex and superior colliculus in mediation of visually guided behavior in the cat. *Science*, *153*, 1544–1547.

Staubli, U., Vanderklish, P. W., & Lynch, G. (1990). An inhibitor of integrin receptors blocks LTP. *Behav. Neural Biol.*, *53*, 1–5.

Strens, L. H., Fogelson, N., Shanahan, P., Rothwell, J. C., & Brown, P. (2003). The ipsilateral human motor cortex can functionally compensate for acute contralateral motor cortex dysfunction. *Curr. Biol.*, *13*(14), 1201–1205.

Syken, J., Grandpre, T., Kanold, P. O., & Shatz, C. J. (2006). PirB restricts ocular-dominance plasticity in visual cortex. *Science*, *313*, 1795–1800.

Valero-Cabre, A., Payne, B. R., & Pascual-Leone, A. (2007). Opposite impact on (14)C-2-deoxyglucose brain metabolism following patterns of high and low frequency repetitive transcranial magnetic stimulation in the posterior parietal cortex. *Exp. Brain Res.*, *176*(4), 603–615.

Valero-Cabre, A., Payne, B. R., Rushmore, J., Lomber, S. G., & Pascual-Leone, A. (2005). Impact of repetitive transcranial magnetic stimulation of the parietal cortex on metabolic brain activity: A 14C-2DG tracing study in the cat. *Exp. Brain Res.*, *163*(1), 1–12.

Vines, B. W., Nair, D. G., & Schlaug, G. (2006). Contralateral and ipsilateral motor effects after transcranial direct current stimulation. *Neuroreport*, *17*(6), 671–674.

Vuilleumier, P., Hester, D., Assal, G., & Regli, F. (1996). Unilateral spatial neglect recovery after sequential strokes. *Neurology*, *46*, 184–189.

Ward, N. S., & Cohen, L. G. (2004). Mechanisms underlying recovery of motor function after stroke. *Arch. Neurol.*, *61*(12), 1844–1848.

Werhahn, K. J., Conforto, A. B., Kadom, N., Hallett, M., & Cohen, L. G. (2003). Contribution of the ipsilateral motor cortex to recovery after chronic stroke. *Ann. Neurol.*, *54*(4), 464–472.

Zhuo, M. (2008). Cortical excitation and chronic pain. *Trends Neurosci.*, *31*(4), 199–207.

10 Exercising Your Brain: Training-Related Brain Plasticity

DAPHNE BAVELIER, C. SHAWN GREEN, AND MATTHEW W. G. DYE

ABSTRACT Learning and brain plasticity are fundamental properties of the nervous system, and they hold considerable promise when it comes to learning a second language faster, maintaining our perceptual and cognitive skills as we age, or recovering lost functions after brain injury. Learning is critically dependent on experience and the environment that the learner has to face. A central question then concerns the types of experience that favor learning and brain plasticity. Existing research identifies three main challenges in the field. First, not all improvements in performance are durable enough to be relevant. Second, the conditions that optimize learning during the acquisition phase are not necessarily those that optimize retention. Third, learning is typically highly specific, showing little transfer from the trained task to even closely related tasks. Against these limiting factors, the emergence of complex learning environments provides promising new avenues when it comes to optimizing learning in real-world settings.

The ability to learn is fundamentally important to the survival of all animals. Brain plasticity, together with the learning it enables, therefore embodies a pivotal evolutionary force. The human species appears remarkable in this respect, as more than a century of research has demonstrated that humans possess the ability to acquire virtually any skill given appropriate training. Yet, while the exceptional capacity of humans to learn should certainly reassure those seeking to design educational or rehabilitative training programs, there are still several key obstacles that need to be overcome before these programs can reach their full potential.

The first is that brain plasticity is typically highly specific. While individuals trained on a task will improve on that very task, other tasks, even closely related ones, often show little or no improvement. Obviously, this obstacle potentially limits the benefits of learning-based interventions, be they educational or clinical. After all, it is of little use to improve the performance of a stroke patient on a visual motion task in the laboratory if this same training will not allow her to effectively see moving cars as she tries to safely cross the street.

DAPHNE BAVELIER and MATTHEW W. G. DYE Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York

C. SHAWN GREEN Department of Psychology, University of Minnesota, Minneapolis, Minnesota

The second obstacle is that while brain plasticity is typically adaptive and beneficial, it can also be maladaptive, dramatically so at times, as when expert string musicians suffer from dystonia or motor weaknesses in their fingers as a result of extensive practice with their instruments.

Finally, and subsumed in the first two obstacles, is the fact that we are still missing the recipe for successful brain plasticity intervention at the practical level. Our current understanding of the causal relationship between one type of training experience and the functional changes it induces through brain plasticity is still very much incomplete.

However, progress is being made in each of these areas. In particular, research in recent years has revealed the potential benefits of what are sometimes termed complex learning environments. These appear to promote behaviorally beneficial plastic changes at a more general level than previously seen. This chapter provides an overview of these recent advances.

Specificity of learning

In the field of learning, transfer of learning from the trained task to even other very similar tasks is generally the exception rather than the rule. This fact is well documented in the field "perceptual learning" literature. For instance, Fiorentini and Berardi (1980) trained subjects to discriminate between two complex gratings that differed only in the relative spatial phase of the two component sinusoids (figure 10.1*A*). Performance on this task improved very rapidly over the course of a single training session and remained consistently high when subjects were tested on two subsequent days. However, when the gratings were rotated by 90 degrees or the spatial frequency was doubled, no evidence of transfer was observed (figure 10.1*B*). Specificity has also been demonstrated in the discrimination of oriented texture objects, where learning is specific to the location and orientation of the trained stimuli (Karni & Sagi, 1991), in the discrimination of dot motion direction, where the learning is specific to the direction and speed of the trained stimuli (Ball & Sekuler, 1982; Saffell & Matthews, 2003), and in some types of hyperacuity tasks, where in addition to being specific for location and orientation, learning can even be specific for the trained eye (Fahle, 2004).

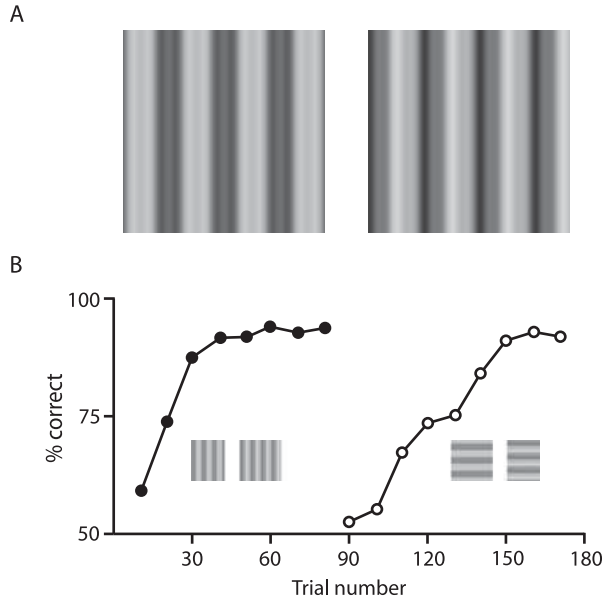


FIGURE 10.1 (*A*) Schematic illustration of the stimulus gratings to be discriminated in Fiorentini and Berardi (1980). (*B*) Subjects' performance on the vertical gratings improved steadily as training proceeded. Yet, when the gratings were abruptly rotated by 90 degrees halfway through the session, performance dropped back to pretraining levels, illustrating the high specificity of the learning.

Similar examples of specificity can also be found in the motor domain (Bachman, 1961; Rieser, Pick, Ashmead, & Garing, 1995). For example, participants trained to aim at a target with their aiming hand visible demonstrate increases in the speed and accuracy of their aiming movements. However, these improvements do not transfer to conditions in which the aiming hand is not visible (Proteau, 1992). In prism adaptation studies, subjects wear goggles that displace the visual world laterally, thus requiring a recalibration of the motor system to bring it back in alignment with the nondisplaced real world. In this literature there is evidence for learning that is specific to the trained limb (Martin, Goodkin, Bastian, & Thach, 1996), to the start and end position of the learned movement, and to the action performed (Redding, Rossetti, & Wallace, 2005; Redding & Wallace, 2006).

Specificity of learning is also a feature of more cognitive learning. For instance, Pashler and Baylis (1991) trained subjects to associate one of three keys with visually presented symbols (left key = P or 2, middle key = V or 8, right key = K or 7). Over the course of multiple training blocks, participant reaction time decreased significantly. However, when new symbols were added that needed to be mapped to the same keys in addition to the learned symbols (left key = P, 2, F, 9; middle key = V, 8, D, 3; right key = K, 7, J, 4), no evidence of transfer was evident. In fact, reaction times to the previously learned symbols increased to pretraining levels. Similarly, studies of object recognition point to highly specific learning. Furmanski and Engel (2000) trained subjects to name backward masked images of common objects over 5 days. Recognition thresholds decreased by up to 20%; however, little transfer was seen when a new set of objects was used. Thus learning did not proceed through general enhancement of vision or by learning the visual context in which the objects were presented but rather at an object-specific level.

Specificity of learning is not just a feature of training-induced brain plasticity. Plasticity as a result of altered experience, even early in life, also leads to surprisingly specific functional changes. For example, individuals born deaf do not exhibit a general enhancement of vision; they exhibit comparable performance to hearing individuals on a range of visual psychophysical thresholds, be it for brightness discrimination, visual flicker, different aspects of contrast sensitivity, or direction and velocity of motion (Bosworth & Dobkins, 2002; Brozinsky & Bavelier, 2004; Finney & Dobkins, 2001). Instead, enhanced performance has been reported only under specific conditions, such as processing of the visual periphery or motion processing, and mainly under conditions of attention. A review of the literature indicates that the changes documented after early deafness are best captured in terms of a change in the spatial distribution of visual spatial attention, whereby deaf individuals exhibit enhanced peripheral attention compared to hearing individuals, with little to no changes in other aspects of vision or visuospatial attention (Bavelier, Dye, & Hauser, 2006).

Enhanced performance through practice: Is it always learning?

Establishing the presence of experience-dependent learning effects is not always straightforward. At least two main classes of effects may masquerade as experience-dependent learning effects—transient effects and effects caused by hidden or unmeasured variables.

Many types of transient effects may indeed be causally related to the training intervention; however, they are not considered true learning effects because they last for only a few minutes following the cessation of training. An excellent example is the so-called Mozart effect, wherein listening to only 10 minutes of a Mozart sonata was reported to lead to significant performance increases on the Stanford Binet IQ spatial-reasoning task (Rauscher, Shaw, & Ky, 1993). Unfortunately, in addition to proving difficult to replicate consistently (Fudin & Lembessis, 2004; McCutcheon, 2000; Rauscher et al., 1997; Steele, Brown, & Stoecker, 1999), the validity of this enhancement as a true learning effect has been questioned, as any positive effects last only a few minutes. The source of the effect has instead been attributed to short-term arousal or mood changes, as several studies have indicated that the type of music further

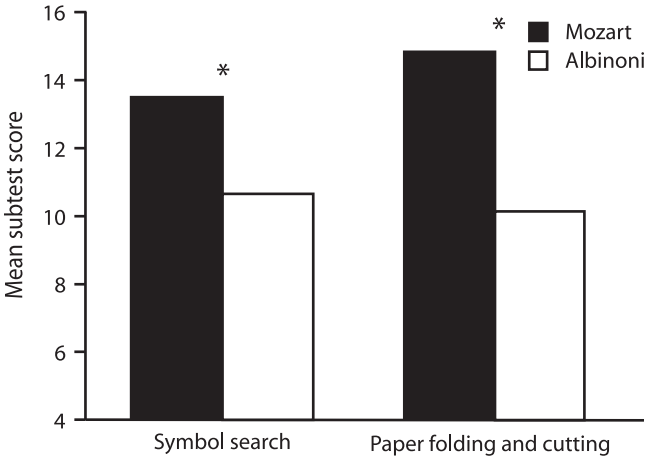


FIGURE 10.2 Participants' performance on the letter-number sequencing test (a measure of working memory skills) and the paper folding and cutting test (a measure of visuospatial constructive skills). Participants were tested shortly after listening to either an up-tempo sonata of Mozart in a major key, which conveyed a mood of happiness, or a slow-tempo adagio of Albinoni in a minor key, which conveyed a mood of sadness. Participants performed better on both tests after listening to the Mozart piece compared to the Albinoni piece. This work illustrates that the "Mozart effect" has little to do with learning per se. Rather, music listening seems to affect performance for better or for worse on a wide variety of tests by changing arousal and mood just before testing. Asterisks denote statistical significance. (Adapted from Schellenberg, Nakata, Hunter, & Tomato, 2007, figure 2; Thompson, Schellenberg, & Husain, 2001, figure 1.)

influences performance. For example, pop music such as "Country House" by Blur led to a greater spatial IQ enhancement than a piece by Mozart (Schellenberg & Hallam, 2005). Further confirming the arousal-mood hypothesis, listening to a high-tempo piece by Mozart was found to lead to better verbal IQ measures than listening to a slower piece by Albinoni (figure 10.2; Schellenberg, Nakata, Hunter, & Tomato, 2007).

Along the same line, studies that have examined the impact of playing violent video games on aggressive behavior may suffer from the same weakness, as the tests used to assess changes in the dependent variables of interest (behavior, cognition, affect, etc.) are typically given within minutes of the end of exposure to the violent video games. Given that violent video games are known to trigger a host of transient physiological changes associated with increased arousal and stress (i.e., "fight-or-flight" responses), it is important to demonstrate that any changes in behavior or cognition are not likewise transient in nature. It is interesting to note that while several recent papers in this field have reported changes in aggressive cognition and affect as well as desensitization to violence immediately following 30 minutes of exposure to violent video games, the same studies failed to find a significant relationship between these variables and being a regular player of violent video games, suggesting that the effects may indeed be fleeting rather than constituting true learned aggression effects (Carnagey & Anderson, 2005; Carnagey, Anderson, & Bushman, 2007).

The second class of effects that may masquerade as experience-dependent learning consists of effects caused by hidden or unmeasured variables that are unrelated to the experience of interest. While these effects may represent learning, they do not represent experience-dependent learning. For instance, it is well documented that individuals who have an active interest taken in their performance tend to improve more than individuals who have no such interest taken—an effect often dubbed the Hawthorne effect (Lied & Karzandjian, 1998). This effect can lead to powerful improvements in performance that have little to do with the specific cognitive training regimen being studied, but instead reflect social and motivational factors that influence performance. In the same vein, the mere presence of mental or physical stimulation may lead to performance changes in groups that are chronically understimulated (as may be the case with the institutionalized elderly), which again would not be considered experience-dependent learning as it is not dependent on the type of experience.

A related issue arises when researchers attempt to infer the presence of experience-dependent learning by examining behavioral differences in groups that perform various activities as part of their everyday lives (for instance, athletes, musicians, or video game players). The obvious concern here is population bias—in other words, inherent differences in abilities may lead to the differences in the activities experienced, rather than the other way around. For example, individuals born with superior hand-eye coordination may be quite successful at baseball and thus preferentially tend to play baseball, while individuals born with poor hand-eye coordination may tend to avoid playing baseball. A hypothetical study that examined differences in hand-eye coordination between baseball players and nonplayers may observe a difference in hand-eye coordination, but it would be erroneous to link baseball experience to superior hand-eye coordination when a population bias was truly at the root of the effect.

Training studies aiming to establish experience-dependent learning should therefore demonstrate (1) benefits that go beyond the temporary arousal or mood changes an experience can induce, and (2) a clear causal link between the specific training experience and learning. The effect of training should be measured at least a full day after completion of training to ensure that it is a robust learning effect. As illustrated by the Mozart effect, training participants for 20 minutes and immediately showing changes in measures of their performance does not mean that a long-lasting alteration of performance has taken place. Furthermore, to establish a definitive causal link between a given form of experience and any enhancement in skills, it is necessary not only to train nonexperts on the experience in question and

to observe the effects of this training, but also to control for the source of this improvement. Training studies should include a group that controls for test-retest effects (i.e., how much improvement can be expected simply from repeating a test) and, just as importantly, for psychological and motivational effects. Control groups that are passive, that are only pre- and posttested but not asked to train or only asked to train at a low level of difficulty on the same task as the experimental group, may not be ideal as such studies fail to differentiate the contribution of motivational factors such as being challenged by the training episode versus true cognitive exercising on performance changes. Finally, evaluation of the efficiency of training critically depends on the choice of outcome measures. Outcome measures closely related to the training experience are more likely to show robust improvements given the specificity of learning discussed earlier. Yet it is critical to show transfer to new tasks within the same domain if one is interested in enhancing skills in a cognitive domain rather than performance on a given specific laboratory test. For example, training on a version of a Stroop task is likely to result in reduced Stroop interference. To what extent does this improvement reflect a generalized improvement in executive skills? Various kinds of transfer and generalization tests that also measure executive skills, but do not do so in the same context or using the same stimuli as the Stroop task, would have to be evaluated before concluding that the training regimen leads to be an improvement in executive skills (see Schmidt & Bjork, 1992, for an excellent review of this issue).

Complex learning environments and general learning

Against a backdrop of highly specific learning, a few training regimens have recently come under close scrutiny, as they seem to induce learning that is much more general than previously thought possible. These learning paradigms are typically more complex than simple laboratory manipulations and correspond to real-life experiences such as musical training, athletic training, and action video game playing.

In the musical domain for instance, Schellenberg (2004) assessed the effect of music lessons on IQ. A large sample of children was randomly assigned to one of four groups. Two groups received music training (keyboard or vocal), one control group received drama training, and the final group received no training. The primary measures of interest were scores on the Wechsler Intelligence Scale for Children (WISC-III) before and after training. While IQ scores increased for children in all groups, the largest increases were observed in the two music training groups. This effect held in all but two of the twelve subtests of the full IQ scale, indicating a widespread beneficial effect on cognition (figure 10.3). Rauscher and colleagues (1997) monitored the spatiotemporal reasoning skills of young children (3–4 years old) who were given 6 months of musical keyboard lessons. Significantly larger improvements in spatiotemporal reasoning were noted in the keyboard-trained children than in two control groups—one a computer training and the other a no-training group (see also Hetland, 2000). Finally, it has also been suggested that music training enhances mathematical ability and verbal memory (Gardiner, Fox, Knowles, & Jefferey, 1996; Graziano, Peterson, & Shaw, 1999; Ho, Cheung, & Chan, 2003). These studies demonstrate a causal effect of music playing on a range of cognitive skills during development. Although the motor component of music lessons is likely to be a key factor, it remains unknown whether different musical activities (string playing, keyboard playing, or singing) differ in altering cognition. Similarly, it is not clear whether these differences remain into adulthood and whether they can be induced through music playing in adulthood.

In the athletic domain, Kiomourtzoglou, Kourtessis, Michalopoulou, and Derri (1998) compared athletes with expertise in various sports (basketball, volleyball, and water polo) on a number of measures of perception and cognition. Expert athletes demonstrated enhancements (compared to novices) in skills that are intuitively important to performance in their given sports. Basketball players exhibited superior selective attention and hand-eye coordination, volleyball players outperformed novices at estimating the speed and direction of a moving object, and water polo players had

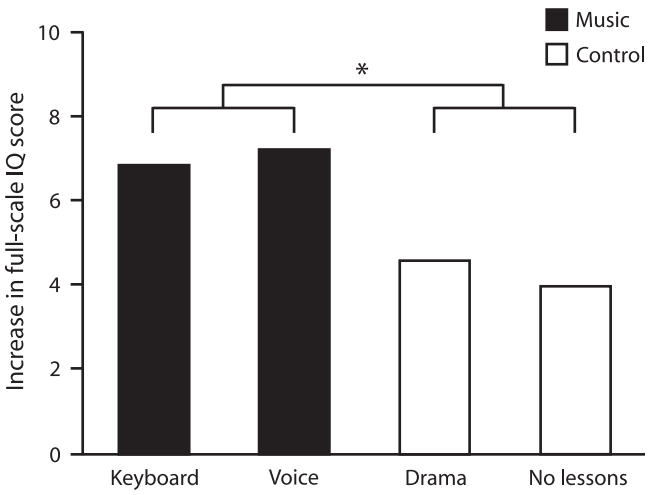


FIGURE 10.3 A large sample of children was randomly assigned to one of four groups. Two groups received music training (keyboard or vocal), a first control group received drama training, and a second control group received no training. The primary measures of interest were scores on the Wechsler Intelligence Scale for Children (WISC-III) obtained before and after training. Children in the music groups showed greater improvements between pre- and post-assessment than the two control groups. This study demonstrates a causal effect of music lessons on a range of cognitive skills during development. Asterisks denote statistical significance. (Data replotted from Schellenberg, 2006.)

faster visual reaction times and better spatial orienting abilities. Lum, Enns, and Pratt (2002), McAuliffe (2004), and Nougier, Azemar, and Stein (1992) observed similar sports-related differences in a Posner cuing task, while Kida, Oda, and Matsumura (2005) demonstrated that trained baseball players respond faster than novices in a go-no-go task (press the button if you see color A, do not press the button if you see color B), but interestingly show no enhancements in simple reaction time tasks (press a button when a light turns on). Unfortunately, no training studies are available at this point to establish a causal link between these performance enhancements and the specific physical activity under investigation. The possibility that aerobic exercise of any sort may enhance cognitive abilities has received much attention lately with respect to aging. Consistently positive results have been reported in many cross-sectional studies comparing older adults who normally exercise with those who do not. Enhancements have been documented in tasks as varied as dual-task performance or executive attention/distractor rejection (for recent reviews see Colcombe et al., 2003; Hillman, Erickson, & Kramer, 2008; Kramer & Erickson, 2007). More training studies are needed to unambiguously establish the causal effect of aerobic exercise on perception and cognition. Yet, taken together, studies of the effect of athletic training and exercise on perception and cognition are tantalizing, and they have prompted renewed interest for demonstrating a causal link between the physical nature of the training regimen and enhancement of cognitive skills.

Perhaps the most popular training regimen over the past decade has been video games. The possibility that perceptual and cognitive abilities are enhanced in video game players has raised much attention (for a review, see Green & Bavelier, 2006b). Indeed, the variety of different skills and the degree to which they are modified in video game players appears remarkable. These include improved hand-eye coordination (Griffith, Voloschin, Gibb, & Bailey, 1983), increased processing in the periphery (Green & Bavelier, 2006c), enhanced mental rotation skills (Sims & Mayer, 2002), greater divided attention abilities (Greenfield, DeWinstanley, Kilpatrick, & Kaye, 1994), faster reaction times (Castel, Pratt, & Drummond, 2005), and even job-specific skills such as laparoscopic manipulation (Rosser, Lynch, Cuddihy, Gentile, & Merrell, 2007) and airplane piloting procedures (Gopher, Weil, & Bareket, 1994). Although intriguing, this literature has little to say about learning per se unless the causal effect of game playing is unambiguously established. So far, only a few studies have established a causal link between video game play and long-lasting changes in performance. Among these is a series of studies that provide compelling evidence that playing action video games—such as first-person perspective shooter games—promotes widespread changes ranging from early sensory functions to higher cognitive functions in adults.

Playing action video games improves fundamental properties of vision (Green & Bavelier, 2007; Li, Polat, Makous, & Bavelier, in press). One visual ability often diminished in patients with poor vision, such as amblyopes or older adults (Bonneh, Sagi, & Polat, 2007), is the ability to read small print, with letters appearing unstable and jumbled. The tendency for the resolvability of letters to be adversely affected by near neighbors, termed crowding, is typically evaluated by asking subjects to identify the orientation of a letter flanked by distractors, and by determining the smallest distance between target and distractors at which subjects can still correctly identify the target (figure 10.4*A*). Individuals with better vision can tolerate distractors being brought nearer to the target while still maintaining high-accuracy performance. To establish the causal effect of action video game playing on this visual skill, a training study was carried out whereby subjects were randomly assigned to one of two training groups: an action video games trained group (e.g., Unreal Tournament,) or a control trained group (e.g., Tetris). Each group was tested pre- and post-training on the crowding task. Participants trained on the action game improved significantly more than those trained on the control game (figure 10.4*B*). The inclusion of a control game group allows us to measure any possible improvements due to test-retest (i.e., familiarity with the task) or to Hawthorne-like effects (Lied & Karzandjian, 1998). Finally, the control games were chosen to be as pleasurable and engrossing as the experimental training games in order to minimize differences in arousal across groups. Critically, posttraining evaluation was always performed at least a day after the completion of the training phase.

Playing action video games was also shown to enhance several different aspects of visual selective attention. Action game training improves the ability of young adults to search their visual environment for a prespecified target, to monitor moving objects in a complex visual scene, and to process a fast-paced stream of visual information (Feng, Spence, & Pratt, 2007; Green & Bavelier, 2003, 2006c, 2006d). In one such experiment, the efficiency with which attention is distributed across the visual field was measured with a visual search task called the Useful Field of View paradigm (Ball, Beard, Roenker, Miller, & Griggs, 1988). This task is akin to looking for a set of keys on a cluttered desk. Subjects are asked to localize a briefly presented peripheral target in a field of distracting objects; accuracy of performance is recorded (figure 10.5*A*). Training on an action video game for just 10 hours improved performance on that task by about 30%, an improvement which is greatly in excess of that which can be induced by training on a control game (figure 10.5*B*). In a related study, Feng and colleagues (2007) showed that performance on the Useful Field of View task differs across gender, with males showing an advantage. Yet, after 10 hours of action game training, this gender difference was

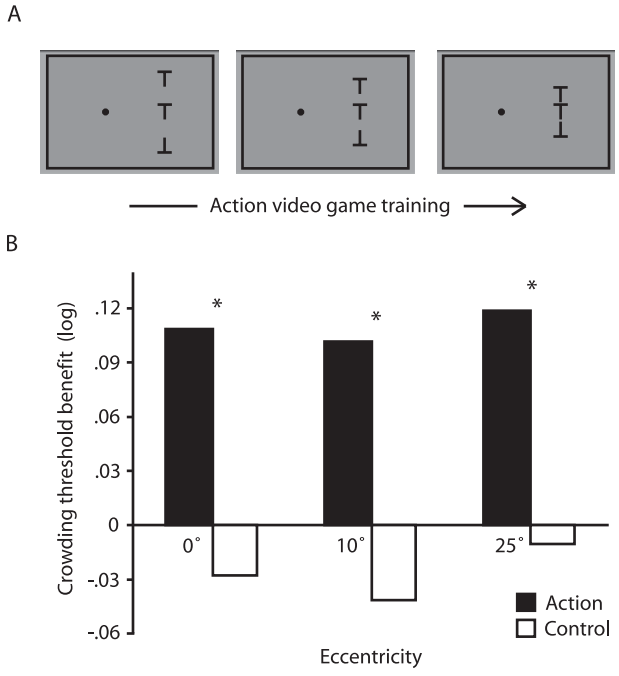


FIGURE 10.4 (*A*) Participants were presented with a display containing three vertically aligned *T*'s and asked to determine whether the central *T* was upright or inverted. Crowding thresholds were measured by determining the smallest distance between the target and the distractors at which participants could still perform this discrimination task with 79% accuracy. Enhanced performance on this task results in participants being able to process more densely packed letters, as illustrated here. Participants were trained either on an action video game hypothesized to enhance their visual resolution or on a control video game. (*B*) The crowding thresholds were measured at three different eccentricities (central vision, 10 degrees, and 25 degrees). This procedure allowed testing of central vision, often thought to have optimal performance, as well as peripheral locations, allowing one to test generalization of learning at untrained locations. The action game training group improved significantly more than the control group at all three eccentricities tested, reflecting generalization of learning at untrained locations and greater plasticity than previously thought in central vision. Asterisks denote statistical significance.

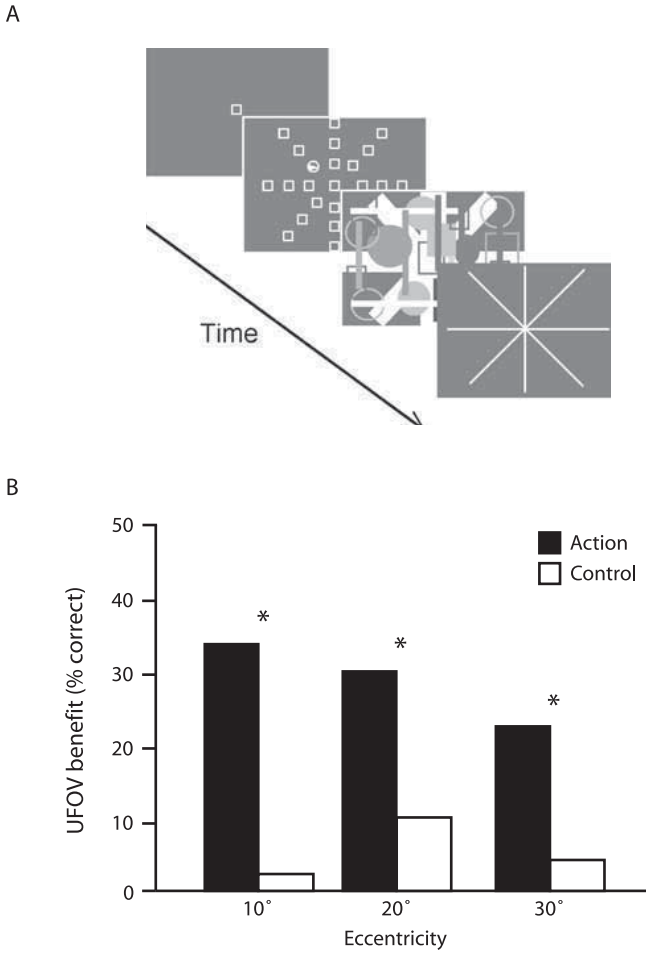


FIGURE 10.5 (*A*) Illustration of the Useful Field of View Task as adapted by Green and Bavelier (2003). Participants viewed a briefly flashed display containing one target, a filled triangle, embedded in a circle of distractors. They were asked to report the location of the target by indicating along which of the main eight directions the target was presented. Half of the participants were trained on an action video game hypothesized to enhance their visual attention, while the other half were trained on a control game. (*B*) Percent correct target localization was measured at each of the three eccentricities tested (10, 20, and 30 degrees of visual angle) before and after training. The action game training group improved more from pre- to post-training tests than the control game training group. This was even the case at 30 degrees of visual angle, an eccentricity seldom used in video gaming, establishing generalization of learning to untrained locations. Asterisks denote statistical significance.

reduced, as well as the oft-documented difference in mental rotation skills between males and females (Feng et al., 2007).

In addition to basic visual skills and selective attention, action game playing has also been linked to better performance on dual tasks (Green & Bavelier, 2006a), task switching (Green & Bavelier, 2006a), and decision-making processes (Green, Pouget, & Bavelier, 2007). A training regimen that promotes robust changes in such a wide range of skills demonstrates that efficient learning transfer can occur given the appropriate training.

Determinants of learning and learning transfer

A major challenge for future work is to pinpoint which factors, or combination of factors, inherent to the complex training regimens discussed earlier are responsible for the enhancement in learning and learning transfer. This point is important both theoretically, in terms of designing models of human learning and behavior, and practically, for those seeking to devise effective rehabilitation programs to ameliorate specific deficits.

The ultimate goal is to see the learner flexibly acquire new knowledge, while using prior knowledge to constrain and accelerate learning. Models of complex human learning,

such as those derived from connectionism or machine learning, provide some clues about the factors that facilitate bottom-up learning based upon the statistics of the input. Recently, the framework of Bayesian inference has been proposed to provide a good first-order model of how subjects learn to optimize behavior in dynamic complex tasks, be they perceptual or cognitive in nature (Courville, Daw, & Touretzky, 2006; Ernst & Banks, 2002; Orbán, Fiser, Aslin, & Lengyel, 2008; Tenenbaum, Griffiths, & Kemp, 2006). Another key feature of recent advances has been the realization that actions and the feedback they provide about the next step to be computed can greatly reduce the computational load of a task, as well as facilitate learning and generalization (Ballard, Hayhoe, Pook, & Rao, 1997; Taagten, 2005). Finally, symbolic cognitive architectures such as SOAR and ACT-R provide insights into how knowledge representations should be structured to explain the acquisition of abstract systems of knowledge, and possibly transfer of knowledge across these systems (Anderson et al., 2004; Lehman, Laird, & Rosenbloom, 1998). Based on this variety of theoretical approaches, one can begin to identify characteristics inherent to complex training regimens that seem more likely to be at the root of general learning. These include, but are not limited to, (1) level of representation, (2) task difficulty, (3) goals, action, and feedback, and (4) motivation and arousal.

Levels of Representation Learning is more likely to be flexible and general if it occurs at the level of richly structured representations that contribute to a wide array of behaviors, rather than if it changes neural networks whose functions are highly specialized. The field of perceptual learning has identified task difficulty as one of the main factors controlling the level of representation at which learning occurs. In their reverse hierarchy theory of perceptual learning, Ahissar and Hochstein (2004) hypothesize that learning is a top-down guided process, where learning occurs at the highest level of representation that is sufficient for the given task. Easy tasks can be learned at a reasonably high level of representation that may be shared with many other tasks, allowing for sizable learning transfer. When tasks become exceedingly difficult—at least in the perceptual domain, such as in Vernier acuity tasks near the hyperacuity range—lower levels of representation with better signal-to-noise ratios are required for adequate task performance. In such cases, only tasks that make use of this low-level neural network, down to the specific retinal location and stimulus orientation, will benefit. Although the reverse hierarchy theory was developed to account for perceptual learning effects, it aligns well with the more general proposal that transfer of learned knowledge to different tasks and contexts will be more likely when learning and inference operate at higher levels of representation.

A key factor in ensuring flexible learning is high variability. Variability is important both at the level of the exemplars to be learned and the context in which they appear (Schmidt & Bjork, 1992). For example, subjects learn to recognize objects in a more flexible way if the objects are presented in a highly variable context (Brady & Kersten, 2003). High contextual variability ensures that subjects learn to ignore the specifics of the objects, such as are brought about by changes in view, lighting, camouflage, or shape, and rather learn to extract more general principles about object category. Statistical approaches such as mutual information show that subjects implicitly develop knowledge of the fragments or chunks that carry information about the categories to be learned (Hegdé, Bart, & Kersten, 2008; Orbán et al., 2008). A key issue then arises as to when these informative fragments allow for learning that generalizes as compared to learning that is item specific. Work on object classification and artificial grammar learning shows that low input variability induces learning at levels of representation that are specific to the items being learned, and thus too rigid to generalize to new stimuli. High variability is crucial in ensuring that the newly learned informative fragments be at levels of representation that can flexibly recombine (Gomez, 2002; Onnis, Monaghan, Christiansen, & Chater, 2004; Reeler, Newport, & Aslin, 2008). Research on the video game Tetris and its effect on mental rotation illustrates this point well. Even though mental rotation is at a premium in Tetris, expert Tetris players have been found to exhibit mental rotation capacities similar to those of naïve subjects, except when tested on Tetris or Tetris-like shapes (Sims & Mayer, 2002). The use of a limited number of shapes in Tetris allows the learner to memorize spatial configurations and moves (Destefano & Gray, 2007). This approach allows for the development of excellent expertise at the game itself, but what is learned in this low-variability game is less likely to generalize to other environments. By this view, an efficient scheme to enforce mental rotation learning would be to use a highly variable set of objects preventing learning of specific configurations.

Task Difficulty The proposal that task difficulty controls the type and rate of learning is implicit in all theories of learning. The perceptual learning literature nicely illustrates the impact of manipulating task difficulty appropriately (Sireteanu & Rettenbach, 1995, 2000). In particular, when it comes to promoting learning transfer, harder tasks are at a disadvantage. For example, in a task where participants had to view arrays of oriented lines and determine which contained a single oddly oriented line, task difficulty was manipulated by limiting exposure time (Ahissar & Hochstein, 1997). With practice, the minimal exposure time that could be tolerated by the participants decreased substantially. Interestingly, when the task was started at a

difficult level (short exposure times), learning was slow and specific for the trained orientation and location. When the task was made easier by starting with long exposure times, learning progressed quickly and transferred to novel orientations. Other conditions that made the task more difficult, such as using small differences in orientation between target and nontarget lines or greater visual eccentricity, also led to the same effect. In the same vein, Liu and Weinshall (2000) demonstrated that learning an easy motion-direction discrimination (9 degrees of motion direction) transferred substantially to novel orientations, whereas Ball and Sekuler (1982) had previously reported no such transfer using the same task but with a greater degree of difficulty (3 degrees of motion direction). Similarly, albeit with barn owls rather than human subjects, Linkenhoker and Knudsen (2002) demonstrated that adult barn owls could adjust to sizable shifts in visual experience (brought about by prism goggles) when the shifts were made in small enough increments. In contrast, large shifts led to no learning in these adult barn owls.

This is not to say that efficient learning will occur through exposure of the learner to situations that are easy to master. In fact, easy tasks that typically require the reenactment of already mastered skills lead to little to no learning (Olesen, Westerberg, & Klingberg, 2004). Intuitively the task difficulty should be set such that the learner can gain some satisfaction from his or her performance. In other words, it should be challenging enough to avoid boredom and lack of interest, but not too hard to allow for sizable positive feedback. This balance may be understood more formally as choosing the task difficulty that allows the learner to optimize over time the amount of reward gained from doing the task. Strikingly, the video game industry may have focused in on the conditions for generalized learning by using variable entry level, and therefore allowing each learner to enter the learning task at the proper level of challenge, and by implementing incremental increases in task difficulty as the game progresses. In our own work on video game training we have acknowledged these principles explicitly by progressing players to the next level of difficulty during training only when they have demonstrated sufficient mastery of their current level. This is not to say that the type of game is unimportant, rather that an appropriate training regimen must also be administered in an appropriate manner.

Goals, Action, and Feedback A productive view of learning holds that it derives from the need to minimize "surprise," or the difference between the anticipated outcome of an event or action and its actual outcome (Courville et al., 2006; Schultz, Dayan, & Montague, 1997). In that framework, actions provide an opportunity for learners to evaluate their internal representations and fine-tune them if a discrepancy is noted between the actual and predicted outcome of the action (Sutton & Barto, 1998). Learning is thus critically under the control of the expected value, or reward, that the learner ascribes to a future event or action and the actual value received as the event or action unfolds.

Although clearly critical for learning, the exact role that feedback plays in learning is a subject of much debate. There are numerous examples demonstrating that feedback is necessary for learning (Herzog & Fahle, 1997; Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006). Yet many counterexamples also exist (Amitay, Irwin, & Moore, 2006; Ball & Sekuler, 1987; Fahle, Edelman, & Poggio, 1995; Karni & Sagi, 1991). The extent to which these are counterexamples is complicated by the fact that, even when experimenter-generated explicit feedback is not provided, if above-threshold stimuli are employed subjects will nevertheless have varying degrees of confidence that their response was correct. Such internally generated confidence judgments could themselves act as feedback signals (Mollon & Danilova, 1996). An added complication stems from the finding that the type of feedback that optimizes learning during performance acquisition is not necessarily that which optimizes learning in the long run (Schmidt & Bjork, 1992). For example, during the learning of a complex arm movement, subjects provided with feedback about their movement-time error after every trial learned faster than those provided with the same feedback but in a summary form every 15 trials. Yet upon retesting two days later, those provided with feedback every 15 trials showed greater accuracy and thus better performance on the task than those provided with feedback every trial (Schmidt, Young, Swinnen, & Shapiro, 1989). Whether feedback frequency systematically affects skill acquisition differently than skill retention remains to be firmly established; yet such findings certainly call for caution in considering the roles of feedback in learning.

While most major theories of learning require that some type of learning signal be present (often in the form of an error signal), they do not necessarily require that the feedback be explicit, nor do they require that feedback be given on a trial-to-trial basis. There are many algorithms that can learn quite efficiently when feedback is only given after a series of actions have been completed (Walsh, Nouri, & Littman, 2007). This is analogous to the situation that commonly occurs in action video games, where feedback (typically in the form of killing an opponent or dying) only becomes available at the conclusion of a very complicated pattern of actions. How best to solve this credit assignment problem, as well as how this affects the generality of what is learned, is a topic of ongoing research (Fu & Anderson, 2008; Ponzi, 2008). Interestingly, complex learning environments with the variety of actions they encompass allow for error signals that are varied both in nature and in time scale, a feature that may facilitate flexible learning.

The importance of reward in learning is already supported by neurophysiology studies that show that the brain systems thought to convey the utility of reward, such as the ventral tegmental area and the nucleus basilis, play a large role in producing plastic changes in sensory areas. In particular, when specific auditory tones are paired with stimulation of either of these structures, the area of primary auditory cortex that represents the given tone increases dramatically in size (Bao, Chan, & Merzenich, 2001; Kilgard & Merzenich, 1998). Interestingly, at least some of the brain areas known to be sensitive to reward have been shown to be extremely active when individuals play action video games. For instance, Koepp and colleagues (1998) demonstrated that roughly the same amount of dopamine is released in the basal ganglia when playing an action video game as when methamphetamines are injected intravenously. Determining the exact role of reward-processing areas in the promotion of learning and neural plasticity will continue to be an area of active research.

Motivation and Arousal Motivation is a critical component of most major theories of learning, with motivation level being posited to depend highly on an individual's internal belief about her ability to meet the current challenge. Vygotsky's (1978) concept of a zone of proximal development matches well with the skill-learning literature discussed previously. According to this theory, motivation is highest and learning is most efficient when tasks are made just slightly more difficult than can be matched by the individual's current ability. Tasks that are much too difficult or much too easy will lead to lower levels of motivation and thus substantially reduced learning. This is not to say that learning will never occur if the task is too difficult or too easy (Amitay et al., 2006; Seitz & Watanabe, 2003; Watanabe, Nanez, & Sasake, 2001), but learning rate should be at a maximum when the task is challenging, yet still doable.

Like motivation, arousal is a key component of many learning theories. The Yerkes-Dodson law predicts that learning is a U-shaped function of arousal level (Yerkes & Dodson, 1908). Training paradigms that lead to low levels of arousal will tend to lead to low amounts of learning, as will training paradigms that lead to excessively high levels of arousal (Frankenhaeuser & Gardell, 1976). Between these extremes there is an arousal level that leads to a maximum amount of learning, which no doubt differs greatly between individuals. Interestingly, video games are known to elicit both the autonomic responses (Hebert, Beland, Dionne-Fournelle, Crete, & Lupien, 2005; Segal & Dietz, 1991; Shosnik, Chatterton, Swisher, & Park, 2000) and neurophysiological responses (Koepp et al., 1998) that are characteristic of arousal. These responses represent a salient difference between traditional learning paradigms and video game play. In the same vein, although again with barn owls, Bergan, Ro, Ro, & Knudsen (2005) observed that adult owls who were forced to hunt (an activity that involves motivation and arousal) while wearing displacing prisms demonstrated significant learning compared to adult owls who wore the prisms for the same period of time, but who were fed dead prey. The latter failed to adapt to the displacing prism.

Conclusions

The field of experience-dependent plasticity is rapidly expanding, thanks in part to new technologies. Cognitive training on handheld devices and job-related training in immersive environments are now within the reach of most institutions, if not individuals. This trend is exciting because the most successful interventions, when it comes to ameliorating deficits in patients or enhancing skills in an educational context, rely on complex training regimens. These regimens require the simultaneous use of perceptual, attentional, memory, and motor skills to trigger learning that goes beyond the specifics of the training regimen itself. New technologies are perfectly positioned to enhance the development of such complex learning environments. For all the excitement, challenges lie ahead. First among these is developing an understanding of which ingredients should be included in training regimens in order to promote widespread learning. Studies of the neural bases of arousal, motivation, and reward processing hold promise in that respect. Second, although the type of improvement desired is usually clear, as when educators or rehabilitation therapists state their goals for a student or a patient, identifying the cognitive component of a training regimen aimed at realizing those goals is not always so straightforward. At first glance, playing action video games does not appear to be a mind-enhancing activity. Yet it seems to generate beneficial effects for perception, attention, and decision making beyond what one may have expected. In contrast, the game Tetris clearly requires mental rotation, and yet it does not lead to a general benefit in mental rotation skill. Cognitive analysis is needed to determine the level of representation at which the learning is most likely to occur given the nature of the training regimen. We are understanding more about the conditions necessary to develop interventions that will lead to generalizable learning effects, and these hold promise for benefiting individuals and the societies within which they live.

acknowledgments This research was supported by grants to DB from the National Institutes of Health (EY016880 and CD04418) and the Office of Naval Research (N00014-07-1-0937). We also thank Bjorn Hubert-Wallander for help in figure preparation and manuscript preparation.

REFERENCES

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*, 401–406.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.*, *8*(10), 457–464.

Amitay, S., Irwin, A., & Moore, D. (2006). Discrimination learning induced by training with identical stimuli. *Nat. Neurosci.*, *9*(11), 1446–1448.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the tmind. *Psychol. Rev.*, *111*(4), 1036–1060.

Bachman, J. C. (1961). Specificity vs. generality in learning and performing two large muscle motor tasks. *Res. Quart.*, *32*, 3–11.

Ball, K., Beard, B., Roenker, D., Miller, R., & Griggs, D. (1988). Age and visual search: Expanding the useful field of view. *J. Opt. Soc. Am. [A]*, *5*(10), 2210–2219.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, *218*, 697–698.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vis. Res.*, *27*, 953–965.

Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behav. Brain Sci.*, *20*(4), 723–742.

Bao, S., Chan, V., & Merzenich, M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature*, *412*, 79–83.

Bavelier, D., Dye, M., & Hauser, P. (2006). Do deaf individuals see better? *Trends Cogn. Sci.*, *10*(11), 512–518.

Bergan, J. F., Ro, P., Ro, D., & Knudsen, E. I. (2005). Hunting increases adaptive auditory map plasticity in adult barn owls. *J. Neurosci.*, *25*(42), 9816–9820.

Bonneh, Y. S., Sagi, D., & Polat, U. (2007). Spatial and temporal crowding in amblyopia. *Vis. Res.*, *47*(14), 1950–1962.

Bosworth, R. G., & Dobkins, K. R. (2002). Visual field asymmetries for motion processing in deaf and hearing signers. *Brain Cogn.*, *49*(1), 170–181.

Brady, M., & Kersten, D. (2003). Bootstrapped learning of novel objects. *J. Vis.*, *3*(6), 413–422.

Brozinsky, C. J., & Bavelier, D. (2004). Motion velocity thresholds in deaf signers: Changes in lateralization but not in overall sensitivity. *Brain Res. Cogn. Brain Res.*, *21*(1), 1–10.

Carnagey, N. L., & Anderson, C. A. (2005). The effects of treward and punishment in violent video games on aggressive affect, cognition, and behavior. *Psychol. Sci.*, *16*(11), 882–889.

Carnagey, N. L., Anderson, C. A., & Bushman, B. J. (2007). The effect of video game violence on physiological desensitization to real-life violence. *J. Exp. Soc. Psychol.*, *43*, 489–496.

Castel, A. D., Pratt, J., & Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychol. (Amst.)*, *119*, 217–230.

Colcombe, A. M., Kramer, A. F., Irwin, D. E., Peterson, M. S., Colcombe, S., & Hahn, S. (2003). Age-related effects of attentional and oculomotor capture by onsets and color singletons as a function of experience. *Acta Psychol. (Amst.)*, *113*, 205–225.

Courville, A., Daw, N., & Touretzky, D. (2006). Bayesian theories of conditioning in a changing world. *Trends Cogn. Sci.*, *10*(7), 294–300.

Destefano, M., & Gray, W. D. (2007). *Use of complimentary actions decreases with expertise*. Paper presented at the Cognitive Science Conference, Nashville.

Ernst, M., & Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Fahle, M. (2004). Perceptual learning: A case for early selection. *J. Vis.*, *4*(10), 879–890.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vis. Res.*, *35*, 3003–3013.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action videogame reduces gender differences in spatial cognition. *Psychol. Sci.*, *18*(10), 850–855.

Finney, E. M., & Dobkins, K. R. (2001). Visual contrast sensitivity in deaf versus hearing populations: Exploring the perceptual consequences of auditory deprivation and experience with a visual language. *Brain Res. Cogn. Brain Res.*, *11*(1), 171–183.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*, 43–44.

Frankenhaeuser, M., & Gardell, B. (1976). Underload and overload in working life: Outline of a multidisciplinary approach. *J. Hum. Stress*, *2*(3), 35–46.

Fu, W. T., & Anderson, J. R. (2008). Solving the credit assignment problem: Explicit and implicit learning of action sequences with probabilistic outcomes. *Psychol. Res.*, *72*, 321–330.

Fudin, R., & Lembessis, E. (2004). The Mozart effect: Questions about the seminal findings of Rauscher, Shaw, and colleagues. *Percept. Mot. Skills*, *98*, 389–405.

Furmanski, C. S., & Engel, S. (2000). Perceptual learning in object recognition: Object specificity and size invariance. *Vis. Res.*, *40*(5), 473–484.

Gardiner, M. F., Fox, A., Knowles, F., & Jefferey, D. (1996). Learning improved by arts training. *Nature*, *381*, 284.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychol. Sci.*, *13*(5), 431–436.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Hum. Factors*, *36*(3), 387–405.

Graziano, A. B., Peterson, M., & Shaw, G. L. (1999). Enhanced learning of proportional math through music training and spatial-temporal reasoning. *Neurol. Res.*, *21*(2), 139–152.

Green, C. S., & Bavelier, D. (2003). Action video games modify visual selective attention. *Nature*, *423*, 534–537.

Green, C. S., & Bavelier, D. (2006a). *Ability to task-switch in action video game players*. Paper presented at the Visual Sciences Conference, Sarasota, FL.

Green, C. S., & Bavelier, D. (2006b). The cognitive neuroscience of video games. In L. Humphreys & P. Messaris (Eds.), *Digital media: Transformations in human communication*. New York: Peter Lang.

Green, C. S., & Bavelier, D. (2006c). Effects of action video game playing on the spatial distribution of visual selective attention. *J. Exp. Psychol. Hum. Percept. Perform.*, *32*(6), 1465–1478.

Green, C. S., & Bavelier, D. (2006d). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*(1), 217–245.

Green, C. S., & Bavelier, D. (2007). Action video game experience alters the spatial resolution of vision. *Psychol. Sci.*, *18*(1), 88–94.

Green, C. S., Pouget, A., & Bavelier, D. (2007). *Action videogame playing improves Bayesian inference for perceptual decision-making*. Paper presented at the Visual Sciences Conference, Sarasota, FL.

Greenfield, P. M., DeWinstanley, P., Kilpatrick, H., & Kaye, D. (1994). Action video games and informal education: Effects on strategies for dividing visual attention. *J. Appl. Dev. Psychol.*, *15*, 105–123.

Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in eye-hand motor coordination of video-game users and non-users. *Percept. Mot. Skills*, *57*, 155–158.

Hebert, S., Beland, R., Dionne-Fournelle, O., Crete, M., & Lupien, S. J. (2005). Physiological stress response to video-game playing: The contribution of built-in music. *Life Sci.*, *76*, 2371–2380.

Hegdé, J., Bart, E., & Kersten, D. (2008). Fragment-based learning of visual object categories. *Curr. Biol.*, *18*(8), 597–601.

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vis. Res.*, *37*, 2133–2141.

Hetland, L. (2000). Learning to make music enhances spatial reasoning. *J. Aesthetic Educ.*, *34*, 179–238.

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: Exercise effects on brain and cognition. *Nat. Rev. Neurosci.*, *9*, 58–65.

Ho, Y. C., Cheung, M. C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*(3), 439–450.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proc. Natl. Acad. Sci. USA*, *88*(11), 4966–4970.

Kida, N., Oda, S., & Matsumura, M. (2005). Intensive baseball practice improves the go/nogo reaction time, but not the simple reaction time. *Cogn. Brain Res.*, *22*(2), 257–264.

Kilgard, M., & Merzenich, M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science*, *279*, 1714–1718.

Kioumourtzoglou, E., Kourtessis, T., Michalopoulou, M., & Derri, V. (1998). Differences in several perceptual abilities between experts and novices in basketball, volleyball, and water-polo. *Percept. Mot. Skills*, *86*(3, Pt. 1), 899–912.

Koepp, M., Gunn, R., Lawrence, A., Cunningham, V., Dagher, A., Jones, T., et al. (1998). Evidence for striatal dopamine release during a video game. *Nature*, *393*, 266–268.

Kramer, A. F., & Erickson, K. I. (2007). Capitalizing on cortical plasticity: Influence of physical activity on cognition and brain function. *Trends Cogn. Sci.*, *11*(8), 342–348.

Lehman, J., Laird, J., & Rosenbloom, P. (1998). A gentle introduction to soar: An architecture for human cognition. In D. Scarborough & S. Sternberg (Eds.), *Methods, models and conceptual issues* (2nd ed., Vol. 4, pp. 211–254). Boston: MIT Press.

Li, R., Polat, U., Makous, W., & Bavelier, D. (in press). Enhancing the contrast sensitivity function through action video game training. *Nat. Neurosci.*

Lied, T. R., & Karzandjian, V. A. (1998). A Hawthorne strategy: Implications for performance measurement and improvement. *Clin. Performance Qual. Health Care, 1998*(6), 4.

Linkenhoker, B. A., & Knudsen, E. I. (2002). Incremental training increases the plasticity of the auditory space map in adult barn owls. *Nature*, *419*(6904), 293–296.

Liu, Z., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vis. Res.*, *40*(1), 97–109.

Lum, J., Enns, J., & Pratt, J. (2002). Visual orienting in college athletes: Explorations of athlete type and gender. *Res. Q. Exerc. Sport*, *73*(2), 156–167.

Martin, T. A., Keating, J. G., Goodkin, H. P., Bastian, A. J., & Thach, W. T. (1996). Throwing while looking through prisms. II. Specificity and storage of multiple gaze-throw calibrations. *Brain*, *119*(Pt. 4), 1199–1211.

McAuliffe, J. (2004). Differences in attentional set between athletes and nonathletes. *J. Gen. Psychol.*, *131*(4), 426–437.

McCutcheon, L. E. (2000). Another failure to generalize the Mozart effect. *Psychol. Rep.*, *87*, 325–330.

Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vis.*, *10*(1), 51–58.

Nougier, V., Azemar, G., & Stein, J. (1992). Covert orienting to central visual cues and sport practice relations in the development of visual attention. *J. Exp. Child Psychol.*, *54*, 315–333.

Olesen, P., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nat. Neurosci.*, *7*(1), 75–79.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). *Variability is the spice of learning, and a crucial ingredient for detecting and generalising in nonadjacent dependencies*. In Proceedings of the Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum.

Orbán, G., Fiser, J., Aslin, R., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proc. Natl. Acad. Sci. USA*, *105*(7), 2745–2750.

Pashler, H., & Baylis, G. (1991). Procedural learning. 2. Intertrial repetition effects in speeded choice tasks. *J. Exp. Psychol. Learn. Mem. Cogn.*, *17*, 33–48.

Ponzi, A. (2008). Dynamical model of salience gated working memory, action selection and reinforcement based on basal ganglia and dopamine feedback. *Neural Net.*, *21*(2–3), 322–330.

Proteau, L. (1992). On the specificity of learning and the role of visual information for movement control. In L. Proteau & D. Elliott (Eds.), *Vision and motor control* (Vol. 85, pp. 67–103). Amsterdam: North Holland.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature*, *365*(6447), 611.

Rauscher, F. H., Shaw, G. L., Levine, L. J., Wright, E. L., Dennis, W. R., & Newcomb, R. L. (1997). Music training causes long-term enhancement of preschool children's spatial-temporal reasoning. *Neurol. Res.*, *19*(1), 2–8.

Redding, G. M., Rossetti, Y., & Wallace, B. (2005). Applications of prism adaptation: A tutorial in theory and method. *Neurosci. Biobehav. Rev.*, *29*(3), 431–444.

Redding, G. M., & Wallace, B. (2006). Generalization of prism adaptation. *J. Exp. Psychol. Hum. Percept. Perform.*, *32*(4), 1006–1022.

Reeler, P., Newport, E. L., & Aslin, R. N. (2008). *The role of distributional information in linguistic categories*. Paper presented at the Boston University Conference on Language Development, Boston.

Rieser, J. J., Pick, H. L., Jr., Ashmead, D. H., & Garing, A. E. (1995). Calibration of human locomotion and models of perceptual-motor organization. *J. Exp. Psychol. Hum. Percept. Perform.*, *21*(3), 480–497.

Rosser, J. C. Jr., Lynch, P. J., Cuddihy, L., Gentile, D. A., Klonsky, J., & Merrell, R. (2007). The impact of video games on training surgeons in the 21st century. *Arch. Surg.*, *142*(2), 181–186.

Saffell, T., & Matthews, N. (2003). Task-specific perceptual learning on speed and direction discrimination. *Vis. Res.*, *43*(12), 1365–1374.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychol. Sci.*, *15*(8), 511–514.

Schellenberg, E. G. (2006). Exposure to music: The truth about the consequences. In G. McPherson (Ed.), *The child as musician: A handbook of musical development*. Oxford, UK: Oxford University Press.

Schellenberg, E. G., & Hallam, S. (2005). Music listening and cognitive abilities in 10 and 11 year olds: The blur effect. *Ann. NY Acad. Sci.*, *1060*, 202–209.

Schellenberg, E. G., Nakata, T., Hunter, P. G., & Tomato, S. (2007). Exposure to music and cognitive performance: Tests of children and adults. *Psychol. Music*, *35*(1), 5–19.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychol. Sci.*, *3*(4), 207–217.

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C., (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.*, *15*, 352–359.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Segal, K. R., & Dietz, W. H. (1991). Physiological responses to playing a video game. *Am. J. Dis. Child.*, *145*(9), 1034–1036.

Seitz, A. R., Nanez, J. E., Sr., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *J. Vis.*, *6*(9), 966–973.

Seitz, A. R., & Watanabe, T. (2003). Psychophysics: Is subliminal learning really passive? *Nature*, *422*, 36.

Shosnik, P. D., Chatterton, R. T., Jr., Swisher, T., & Park, S. (2000). Modulation of attentional inhibition by norepinephrine and cortisol after psychological stress. *Int. J. Psychophysiol.*, *36*(1), 59–68.

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Appl. Cogn. Psychol.*, *16*, 97–115.

Sireteanu, R., & Rettenbach, R. (1995). Perceptual learning in visual search: Fast, enduring but non-specific. *Vis. Res.*, *35*, 2037–2043.

Sireteanu, R., & Rettenbach, R. (2000). Perceptual learning in visual search generalizes over tasks, locations, and eyes. *Vis. Res.*, *40*, 2925–2949.

Steele, K. M., Brown, J. D., & Stoecker, J. A. (1999). Failure to confirm the Rauscher and Shaw description of recovery of the Mozart effect. *Percept. Mot. Skills*, *88*, 843–848.

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Taagten, N. A. (2005). Modeling parallelization and flexibility improvements in skill acquisition: From dual tasks to complex dynamic skills. *Cogn. Sci.*, *29*(3), 421–455.

Tenenbaum, J., Griffiths, T., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn. Sci.*, *10*, 309–318.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychol. Sci.*, *12*, 248–251.

Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Walsh, T. J., Nouri, A., Li, L., & Littman, M. L. (2007). Planning and learning in environments with delayed feedback. In *Lecture notes in computer science* (Vol. 4701, pp. 442–453). Berlin: Springer.

Watanabe, T., Nanez, J., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*, 844–848.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.*, *18*, 459–482.

11 Profiles of Development and Plasticity in Human Neurocognition

COURTNEY STEVENS AND HELEN NEVILLE

ABSTRACT We describe changes in neural organization and related aspects of processing after naturally occurring alterations in auditory, visual, and language experience. The results highlight the considerable differences in the degree and time periods of neuroplasticity displayed by different subsystems within vision, hearing, language, and attention. We also describe results showing the two sides of neuroplasticity, that is, the capability for enhancement and the vulnerability to deficit. Finally we describe several intervention studies in which we have targeted systems that display more neuroplasticity and show significant improvements in cognitive function and related aspects of brain organization.

Extensive research on animals has elucidated both genetic and environmental factors that constrain and shape neuroplasticity (Hunt et al., 2005; Garel, Huffman, & Rubenstein, 2003; Bishop et al., 1999; Bishop, 2003). Such research, together with noninvasive neuroimaging and genetic sequencing techniques, has guided a burgeoning literature characterizing the nature, time course, and mechanisms of neuroplasticity in humans (Pascual-Leone, Amedi, Fregni, & Merabet, 2005; Bavelier & Neville, 2002; Movshon & Blakemore, 1974). Electron microscopic studies of synapses and neuroimaging studies of metabolism and of gray and white matter development in the human brain reveal a generally prolonged postnatal development that nonetheless displays considerable regional variability in time course (Chugani, Phelps, & Mazziotta, 1987; Huttenlocher & Dabholkar, 1997; Neville, 1998; Webb, Monk, & Nelson, 2001). In general, development across brain regions follows a hierarchical progression in which primary sensory areas mature before parietal, prefrontal, and association regions important for higher-order cognition (Giedd et al., 1999; Gogtay et al., 2004). Within each region there is a pattern of prominent overproduction of synapses, dendrites, and gray matter that is subsequently pruned back to about 50% of the maximum value, which is reached at different ages in different regions. The prolonged developmental time course and considerable pruning of connections are considered major forces that permit and constrain human neuroplasticity. Recently an additional factor that appears to be important has been identified. The occurrence of polymorphisms in some genes is widespread in humans and rhesus monkeys but apparently not in other primate species. Polymorphisms provide the capability for environmental modification of the effects of gene expression (gene × environment interactions), and such effects have been observed in rhesus monkeys and humans (Suomi, 2003, 2004, 2006; Sheese, Voelker, Rothbart, & Posner, 2007; Bakermans-Kranenberg, Van Ijzendoom, Pijlman, Mesman, & Femmie, 2008).

For several years we have employed psychophysics, electrophysiological (ERP), and magnetic resonance imaging (MRI) techniques to study the development and plasticity of the human brain. We have studied deaf and blind individuals, people who learned their first or second spoken or signed language at different ages, and children of different ages and of different cognitive capabilities. As detailed in the sections that follow, in each of the brain systems examined in this research—including those important in vision, audition, language, and attention—we observe the following characteristics:

- Different brain systems and subsystems and related sensory and cognitive abilities display different degrees and time periods ("profiles") of neuroplasticity. These may depend on the variable time periods of development and redundant connectivity displayed by different brain regions.
- Neuroplasticity within a system acts as a double-edged sword, conferring the possibility for either enhancement or deficit.
- Multiple mechanisms both support and constrain modifiability across different brain systems and subsystems.

In the sections that follow, we describe our research on neuroplasticity within vision, audition, language, and attention. In each section, we note different profiles of plasticity observed in the system, situations in which enhancements

COURTNEY STEVENS Willamette University, Salem, Oregon
HELEN NEVILLE Department of Psychology, University of Oregon, Eugene, Oregon

versus deficits are observed, and likely mechanisms contributing to these different profiles of plasticity. A final section describes our preliminary studies testing the hypothesis, raised by this basic research on human neuroplasticity, that interventions that target the most plastic, and thus potentially vulnerable, neurocognitive systems can protect and enhance children with, or at risk for, developmental deficits.

Vision

In a number of studies we observe that some, but not all, aspects of visual function are enhanced in deaf adults. Those aspects of vision showing the greatest changes are mediated by structures along the dorsal visual pathway that have been shown to be important in the representation of the peripheral visual fields, as well as in motion processing. By contrast, aspects of processing mediated by the ventral visual pathway, including color perception and processing within the central visual field, are not altered (Baizer, Ungerleider, & Desimone, 1991; Bavelier et al., 2001; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Livingstone & Hubel, 1988; Merigan, 1989; Merigan & Maunsell, 1990; Schiller & Malpeli, 1978; Ungerleider & Mishkin, 1982; Ungerleider & Haxby, 1994; Zeki et al., 1991). For example, congenitally deaf individuals have superior motion detection compared with hearing individuals for peripheral, but not central, visual stimuli (Neville, Schmidt, & Kutas, 1983; Neville & Lawson, 1987b; Stevens & Neville, 2006). These behavioral improvements are accompanied by increases in the amplitudes of early event-related potentials (ERPs) and increased functional magnetic resonance imaging (fMRI) activation in motion-sensitive middle temporal (MT) and middle superior temporal (MST) areas of the dorsal visual pathway (Bavelier et al., 2000, 2001; Neville et al., 1983; Neville & Lawson, 1987b).

In a study comparing ERPs to isoluminant color stimuli (designed to activate the ventral pathway) and motion stimuli (designed to activate the dorsal pathway), no differences were observed between hearing and congenitally deaf individuals in ERPs to color stimuli. In contrast, ERPs to motion were significantly larger and distributed more anteriorly in deaf than in hearing subjects. These differences were only observed for stimuli presented in the peripheral visual field (Armstrong, Hillyard, Neville, & Mitchell, 2002). These results are consistent with the hypothesis that early auditory deprivation has more pronounced effects on the functions of the dorsal than the ventral visual pathway. A parallel literature on developmental disorders suggests that the dorsal visual pathway might also be more vulnerable to deficit in certain developmental disorders, including autism, Williams and fragile X syndromes, and reading or language impairments (Atkinson, 1992; Atkinson et al., 1997; Eden et al., 1996). For example, a number of studies indicate that at least some individuals with specific reading disorder, or dyslexia, have lower sensitivity to detecting coherent motion in random-dot kinetograms despite showing normal thresholds for detecting coherent form in similar arrays of static line segments (Cornelissen, Richardson, Mason, Fowler, & Stein, 1995; Everatt, Bradshaw, & Hibbard, 1999; Hansen, Stein, Orde, Winter, & Talcott, 2001; Talcott, Hansen, Assoku, & Stein, 2000). Dyslexic individuals also show higher thresholds for detecting changes in the speed of motion flow fields (Demb, Boynton, Best, & Heeger, 1998), as well as higher critical flicker fusion thresholds for monochromatic, but not isoluminant, color stimuli when tested with a paradigm using identical task structure to assess each visual pathway (Sperling, Lu, Manis, & Seidenberg, 2003). In addition, there are reports that dyslexic individuals show deficits in pattern contrast sensitivity for high-contrast, low-spatial frequency gratings (Lovegrove, Martin, & Slaghuis, 1986). The behavioral evidence for a visual deficit in dyslexia has been corroborated by recent neuroimaging studies showing decreased (Demb et al., 1998) or even nonsignificant (Eden et al., 1996) activations in motion-sensitive areas MT/MST of dyslexic individuals, though no differences are observed during stationary pattern processing (Eden et al.). These results are parallel and opposite to those described previously showing *improved* behavioral performance and *increased* MT/MST activation in response to motion stimuli in congenitally deaf adults.

Taken together, these data suggest that the dorsal visual pathway may exhibit a greater degree of neuroplasticity than the ventral visual pathway, rendering it capable of either enhancement (as is the case following congenital deafness) or deficit (as is the case in some individuals with some developmental disorders). However, the two literatures have developed largely in parallel, and different tasks have been used to assess dorsal and ventral visual pathway function in each literature. To address this limitation, in a recent study we used the same tasks to assess visual function in both dyslexic adults and congenitally deaf adults, as well as matched controls (Stevens & Neville, 2006). We observed that whereas neither deaf nor dyslexic adults differ from matched controls on a central visual field contrast sensitivity task (figure 11.1*A*), on a peripheral motion detection task, deaf adults show enhancements whereas dyslexic adults show deficits on the same task (figure 11.1*B*). These findings help bridge the two literatures and suggest that the dorsal and ventral pathways show different profiles of neuroplasticity.

A number of mechanisms may render the dorsal pathway more developmentally labile, either to enhancement or deficit, including subsystem differences in rate of maturation, extent and timing of redundant connectivity, and presence of chemicals and receptors known to be important in plasticity. For example, anatomical studies suggest that connections within the regions of the visual system that represent

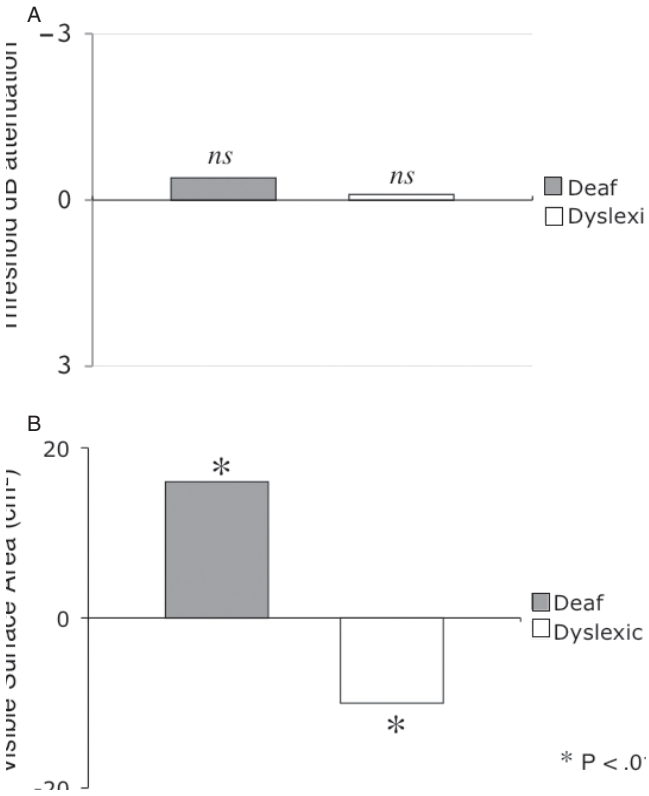


Figure 11.1 Performance on two visual tasks for deaf participants (gray bars) and dyslexic participants (white bars) relative to matched control groups. The zero line represents performance of the respective control groups. (*A*) On a central visual field contrast sensitivity task, neither deaf nor dyslexic participants differed from matched controls. (*B*) On a peripheral visual field motion detection task, deaf participants showed enhancements ($P < .001$) and dyslexic participants showed deficits ($P < .01$) relative to matched controls. (Data from Stevens & Neville, 2006.)

the central visual field are more strongly genetically specified, whereas connections within the portions of the visual system that represent the visual periphery contain redundant connections that can be shaped by experience over a longer developmental time course (Chalupa & Dreher, 1991). A molecular difference has also been observed between the two visual pathways. In cats and monkeys the dorsal pathway has a greater concentration of the Cat-301 antigen, a molecule hypothesized to play a role in stabilizing synaptic connections by means of experience-dependent plasticity (DeYoe, Hockfield, Garren, & Van Essen, 1990; Hockfield, 1983). Moreover, recent anatomical studies in nonhuman primates (Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003) and neuroimaging studies of humans (Eckert et al., 2008) report cross-modal connections between primary auditory cortex and the portion of primary visual cortex that represents the periphery (anterior calcarine sulcus). In addition there is considerable, though not unequivocal, evidence indicating that the dorsal pathway matures more slowly than the ventral pathway (Hickey, 1981; Hollants-Gilhuijs, Ruijter, & Spekreijse, 1998a, 1998b; Packer, Hendrickson, & Curcio, 1990). Further, in developmental studies using the color and motion stimuli described previously and in Armstrong and colleagues (2002), we observed that while children aged 6–19 years show responses to color stimuli that are very similar to adults, their ERPs to the motion stimuli are delayed in latency relative to those for adults (Coch, Skendzel, Grossi, & Neville, 2005; Mitchell & Neville, 2004). Together, these anatomical, chemical, and developmental mechanisms could render the dorsal pathway more modifiable by experience and more likely to display either enhanced or deficient processing.

In addition to enhanced dorsal pathway functioning we have recently observed that deaf (but not hearing) participants recruit a large, additional network of supplementary cortical areas when processing far peripheral relative to central flickering visual stimuli (Scott, Dow, & Neville, 2003, see figure 11.2). These include contralateral primary auditory cortex (figure 11.2). Studies of a mouse model of congenital deafness suggest that altered subcortical-cortical connectivity could account for such changes (Hunt et al., 2005). In deaf but not hearing mice the retina projects to the medial (auditory) geniculate nucleus as well as the lateral (visual) geniculate nucleus. In our study of deaf humans we also observe significant increases in anterior, primary visual cortex and regions associated with multisensory integration (STS), motion processing (MT/MT+), and attention (posterior parietal and anterior cingulate regions) (Dow, Scott, Stevens, & Neville, 2006; Scott et al., 2003; Scott, Dow, Stevens, & Neville, under review). In a separate study, we used structural equation modeling to estimate the strength of cortical connections between early visual areas (V1/V2), area MT/MST, and part of the posterior parietal cortex (PPC) (Bavelier et al., 2000). During attention to the center the connectivity was comparable across groups, but during the attend-periphery condition the effective connectivity between MT/MST and PPC was increased in the deaf as compared with the hearing subjects. The findings of increased activation and effective connectivity between visual areas and areas important in attention suggest that the enhanced responsiveness to peripheral motion in deaf individuals may be in part linked to increases in attention (see next section for further discussion).

Audition

To test whether the specificity of plasticity observed in the visual system generalizes to other sensory systems, we have conducted studies of the effects of visual deprivation on the development of the auditory system. Although less is known about the organization of the auditory system, as in the visual system there are large (magno) cells in the medial geniculate nucleus that conduct faster than the smaller

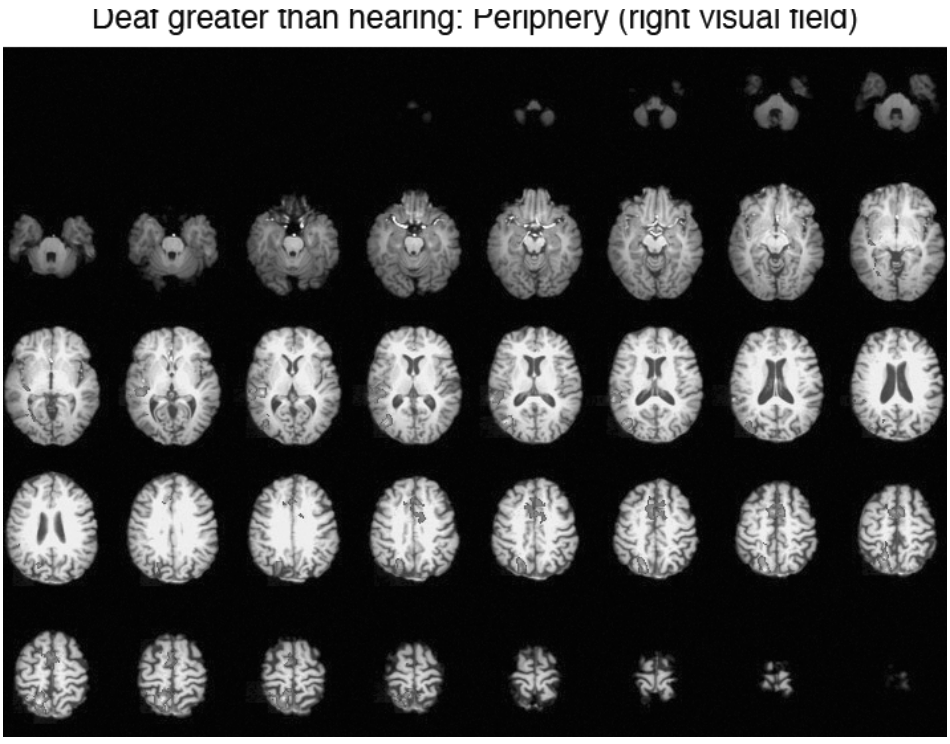


FIGURE 11.2 Deaf and hearing participants completed a visual retinotopy experiment that included mapping of far peripheral visual space. The data show regions where activation was greater in deaf versus hearing participants in response to more peripheral visual stimuli presented in two distinct experiments (45–56° versus 11–23° and 11–15° versus 2–7°). Significant clusters included contralateral auditory cortex, STS, MT, anterior visual cortex, IPS, and anterior cingulate. (See color plate 11.)

(parvo) cells, and recent evidence suggests that there may be dorsal and ventral auditory processing streams with different functional specializations (Rauschecker, 1998). Furthermore, animal and human studies of blindness have reported changes in the parietal cortex (i.e., dorsal pathway) as a result of visual deprivation (Hyvarinen & Linnankoski, 1981; Pascual-Leone et al., 2005; Weeks et al., 2000).

To determine whether similar patterns of plasticity occur following auditory and visual deprivation, we developed an auditory paradigm similar to one of the visual paradigms employed in our studies of deaf adults. Participants detected infrequent pitch changes in a series of tones that were preceded by different interstimulus intervals (Röder, Rösler, Hennighausen, & Näcker, 1996). Congenitally blind participants were faster at detecting the target and displayed ERPs that were less refractory, that is, recovered amplitude faster than normally sighted participants. These results parallel those of our study showing faster amplitude recovery of the visual ERP in deaf than in hearing participants (Neville et al., 1983) and suggest that rapid auditory and visual processing may show specific enhancements following sensory deprivation.

Similar to the two sides of plasticity observed in the dorsal visual pathway, the refractory period for rapidly presented acoustic information, which is enhanced in the blind, shows deficits in many developmental disorders (Bishop & McArthur, 2004; Tallal & Piercy, 1974; Tallal, 1975, 1976). In a study of children with specific language impairment (SLI), we observed that auditory ERPs were smaller (i.e., more refractory) than in controls at short interstimulus intervals (Neville, Coffey, Holcomb, & Tallal, 1993). This finding suggests that in audition, as in vision, neural subsystems that display more neuroplasticity show both greater potential for enhancement and also greater vulnerability to deficit under other conditions.

The mechanisms that give rise to greater modifiability of rapid auditory processing are as yet unknown. However, as mentioned earlier, some changes might be greater for magnocellular layers of the medial geniculate nucleus. For example, magno cells in both the lateral and medial geniculate nucleus are smaller than normal in dyslexia (Galaburda & Livingstone, 1993; Galaburda, Menard, & Rosen, 1994). Rapid auditory processing, including the recovery cycles of neurons, might also engage aspects of attention to a greater degree than other aspects of auditory processing. In the case of congenital blindness, changes in auditory processing may be facilitated by compensatory reorganization. A number of studies have confirmed that visual areas are functionally involved in nonvisual tasks in congenitally blind adults (Cohen, Weeks, Celnik, & Hallett, 1999; Sedato et al., 1996). More recently, studies have reported highly differentiated auditory language processing in primary visual cortex in

congenitally blind humans (Burton et al., 2002; Röder, Stock, Bien, Neville, & Rösler, 2002). Thus aspects of auditory processing that either depend upon or can recruit multimodal, attentional, or normally visual regions may show greater degrees of neuroplasticity. Parallel studies of animals have revealed information about mechanisms underlying this type of change. For example, in blind mole rats, normally transient, weak connections between the ear and primary visual cortex become stabilized and strong (Bavelier & Neville, 2002; Cooper, Herbin, & Nevo, 1993; Doron & Wollberg, 1994; Heil, Bronchti, Wollberg, & Scheich, 1991).

Language

It is reasonable to hypothesize that the same principles that characterize neuroplasticity of sensory systems—including different profiles, degrees, and mechanisms of plasticity—also characterize language. Here, we focus on the subsystems of language examined in our studies of neuroplasticity, including those supporting semantics, syntax, and speech segmentation.

Several ERP and fMRI studies have described the nonidentical neural systems that mediate semantic and syntactic processing. For example, semantic violations in sentences elicit a bilateral negative potential that is largest around 400 ms following the semantic violation (Kutas & Hillyard, 1980; Neville, Nicol, Barss, Forster, & Garrett, 1991; Newman, Ullman, Pancheva, Waligura, & Neville, 2007). In contrast, syntactic violations elicit a biphasic response consisting of an early, left-lateralized anterior negativity (LAN) followed by a later, bilateral positivity, peaking over posterior sites ~600 ms after the violation (P600; Friederici, 2002; Neville et al., 1991). The LAN is hypothesized to index more automatic aspects of the processing of syntactic structure and the P600 to index later, more controlled processing of syntax associated with attempts to recover the meaning of syntactically anomalous sentences. These neurophysiological markers of language processing show a degree of biological invariance as they are also observed when deaf and hearing native signers process American Sign Language (ASL) (Capek, 2004; Capek et al., under review). While spoken and signed language processing share a number of modality-independent neural substrates, there is also specialization based on language modality. The processing of ASL, for example, is associated with additional and/or greater recruitment of right-hemisphere structures, perhaps owing to the use of spatial location and motion in syntactic processing in ASL (Capek et al., 2004; Neville et al., 1998). In support of this hypothesis, we have recently shown that syntactic violations in ASL elicit a more bilateral anterior negativity for violations of spatial syntax, whereas a left-lateralized anterior negativity is observed for other classes of syntactic violations in ASL (Capek et al., under review).

We conducted a series of ERP studies to develop a neural index of one aspect of phonological processing: speech segmentation. By 100 ms after word onset, syllables at the beginning of a word elicit a larger negativity than acoustically similar syllables in the middle of the word (Sanders & Neville, 2003a). This effect has been demonstrated with natural speech and with synthesized nonsense speech in which only newly learned lexical information could be used for segmentation (Sanders, Newport, & Neville, 2002). The early segmentation ERP effect resembles the effect of temporally selective attention, which allows for the preferential processing of information presented at specific time points in rapidly changing streams, and it has also been shown to modulate early (100 ms) auditory ERPs (Lange, Rösler, & Röder, 2003; Lange & Röder, 2005; Sanders & Astheimer, in press). Thus the neural mechanisms of speech segmentation may rely on the deployment of temporally selective attention during speech perception to aid in processing the most relevant rapid acoustic changes.

To the extent that language is made up of distinct neural subsystems, it is possible that, as in vision and audition, these subsystems show different profiles of neuroplasticity. In support of this hypothesis, behavioral studies of language proficiency in second-language learners document that phonology and syntax are particularly vulnerable following delays in second-language acquisition (Johnson & Newport, 1989). In several studies, we have examined whether delays in second-language exposure are also associated with differences in the neural mechanisms underlying these different language subsystems. In one study, we compared the ERP responses to semantic and syntactic errors in English among Chinese/English bilinguals who were first exposed to English at different ages (Weber-Fox & Neville, 1996). Accuracy in judging the grammaticality of the different types of syntactic sentences and their associated ERPs were affected by delays in second-language exposure as short as 4–6 years. By comparison, the N400 response and the behavioral accuracy in detecting semantic anomalies were altered only in subjects who were exposed to English after 11–13 years of age. In studies of the effects of delayed second-language acquisition on indices of speech segmentation, second-language learners who were exposed to their second language late in life (>14 years) show a delay in the ERP measure of speech segmentation when processing their second language (Sanders & Neville, 2003b).

Many deaf children are born to hearing parents and, because of their limited access to the spoken language that surrounds them, do not have full access to a first language until exposed to a signed language, which often occurs very late in development. Behavioral studies of deaf individuals with delayed exposure to sign language indicate that with increasing age of acquisition, proficiency in sign language decreases (Mayberry & Eichen, 1991; Mayberry, 1993;

Mayberry, Lock, & Kazmi, 2002; Mayberry, 2003). Recently, studies have examined the effects of this delayed first-language acquisition on brain organization. We employed fMRI to examine whether congenitally deaf individuals who learned ASL later in life showed a different neural organization for ASL. In this study, we demonstrated that whereas the right angular gyrus is active when native signers process ASL, it is not in individuals who acquired ASL after puberty (Newman, Bavelier, Corina, Jezzard, & Neville, 2002). Employing ERPs, we have also studied groups of deaf individuals who acquired ASL either from birth, from 2 to 10 years, or between 11 and 21 years of age (Capek, 2004; Capek et al., in preparation). In all three groups of participants, the N400 index of semantic processing displays the same amplitude, latency, and cortical distribution. However, the early anterior negativity thought to index more automatic aspects of syntactic processing is only evident in those who acquired ASL before the age of 10 years. These data suggest that, in contrast to semantic processing, aspects of syntactic processing are subject to maturational constraints that render them more vulnerable following delays in either first- or second-language acquisition.

Several lines of evidence suggest that language proficiency might be a key factor in predicting the variability observed in the neural substrates of syntax. For example, we have observed that the neural response to syntactic violations also differs among monolingual native English speakers who vary in language proficiency. Specifically, adults who score lower on standardized tests of grammatical knowledge show a less left-lateralized and more prolonged ERP response to grammatical violations (Pakulak, Hyde, Jackobs, & Neville, 2007; Pakulak & Neville, under review). In developmental studies as well, the neural response to known and unknown words and to syntactic anomalies is more strongly predicted by a child's language proficiency than by chronological age (Adamson, Mills, Appelbaum, & Neville, 1998; Adamson-Harris, Mills, & Neville, 2000; Mills, Coffey-Corina, & Neville, 1993, 1997). Furthermore, the development of neural systems important for syntactic processing shows a longer time course than systems important for semantic processing (Hahne, Eckstein, & Friederici, 2004; Sabourin, Pakulak, Paulsen, Fanning, & Neville, 2007; and unpublished observations from data in our laboratory), again suggesting that systems with a longer developmental time course may be more modifiable during development.

Attention

As noted previously, many of the changes in vision, audition, and language observed in studies of neuroplasticity may depend at least in part on selective attention. The importance of selective attention for certain types of adult neuroplasticity is strongly supported by animal research. For example, when monkeys are provided with extensive exposure to auditory and tactile stimuli, experience-dependent expansions in associated auditory or somatosensory cortical areas occur, but *only* when attention is directed toward those stimuli in order to make behaviorally relevant discriminations (Recanzone, Jenkins, Hradek, & Merzenich, 1992; Recanzone, Schreiner, & Merzenich, 1993). Mere exposure is not enough. These data strongly suggest that attention is important in enabling neuroplasticity. Given this suggestion, as well as the central role of attention in learning more generally, we have conducted several studies on the development and neuroplasticity of attention.

In these studies, we examined the effects of sustained, selective attention on neural processing employing the "Hillyard principle," that is, while keeping the physical stimuli, arousal levels, and task demands constant. For example, competing streams of stimuli are presented (e.g., two different trains of auditory stimuli delivered to different ears), with participants alternating attention to one stream at a time in order to detect rare target events. By comparing neural activity to the same physical stimuli when attended versus ignored, the effects of selective attention can be ascertained. Studies with fMRI revealed that selective attention modulates the magnitude and extent of cortical activation in the relevant processing areas (Corbetta et al., 1990). Complementary studies using the ERP methodology have clarified the time course of attentional modulation. These studies revealed that in adults, selective attention amplifies the sensorineural response by 50–100% during the first 100 ms of processing (Hillyard, Hink, Schwent, & Picton, 1973; Hillyard, Di Russo, & Martinez, 2003; Luck, Woodman, & Vogel, 2000; Mangun & Hillyard, 1990). This early attentional modulation is in part domain general in that it is observed across multiple sensory modalities and in selection based on spatial, temporal, or other stimulus attributes. Moreover, in between-group and change-over-time comparisons, ERPs can separately index processes of signal enhancement (ERP amplitude gains for attended stimuli) and distractor suppression (amplitude reductions for unattended stimuli).

In a number of studies, we have documented that neuroplasticity in the early neural mechanisms of selective attention, as in other neural systems, shows considerable specificity. In the case of adults born deaf, employing ERPs and fMRI, we observed enhancements of attention that were specific to the peripheral, but not central, visual field (Bavelier et al., 2000, 2001; Neville & Lawson, 1987b). In parallel studies of auditory spatial attention among congenitally blind adults, we have observed similar specificity. When attending to central auditory space, blind and sighted participants displayed similar localization abilities and ERP attention effects. In contrast, in the periphery, blind participants were superior to sighted controls at localizing sounds

in peripheral auditory space, and ERPs revealed a sharper tuning of early spatial attention mechanisms (the N1 attention effect) (Röder et al., 1999). In a recent study of adults blinded later in life, we observed possible limits on the time periods during which these early mechanisms of attention are enhanced (Fieger, Röder, Teder-Sälejärvi, Hillyard, & Neville, 2006). Whereas adults blinded later in life showed similar behavioral improvements in peripheral auditory attention, these improvements were mediated by changes in the tuning of later ERP indices of attention, several hundred milliseconds after stimulus onset (i.e., P300). There were no group differences in the early (N1) attention effects. If the early neural mechanisms of selective attention can be enhanced after altered experience, it is possible that, as with other systems that display a high degree of neuroplasticity, attention may be particularly vulnerable during development. In line with this hypothesis, recent behavioral studies suggest that children at risk for school failure, including those with poor language or reading abilities or from lower socioeconomic backgrounds, exhibit deficits in aspects of attention including filtering and noise exclusion (Atkinson, 1991; Cherry, 1981; Farah et al., 2006; Lipina, Martelli, Vuelta, & Colombo, 2005; Noble, Norman, & Farah, 2005; Sperling, Lu, Manis, & Seidenberg, 2005; Stevens, Sanders, Andersson, & Neville, 2006; Ziegler, Pech-Georgel, George, Alanio, & Lorenzi, 2005). These attentional deficits span linguistic and nonlinguistic domains within the auditory and visual modalities, suggesting that the deficits are both domain general and pansensory. In order to determine whether these attentional deficits can be traced to the earliest effects of attention on sensorineural processing, we have recently used ERPs to examine the neural mechanisms of selective attention in typically developing, young children and in groups of children at risk for school failure.

These studies were modeled after those we and others have used with adults (Hillyard et al., 1973; Neville & Lawson, 1987a; Röder et al., 1999; Woods, 1990). The task was designed to be difficult enough to demand focused selective attention, while keeping the physical stimuli, arousal levels, and task demands constant. Two different children's stories were presented concurrently from speakers to the left and right of the participant. Participants were asked to attend to one story and ignore the other. Superimposed on the stories were probe stimuli to which ERPs were recorded. Adults tested with this paradigm showed typical N1 attention effects (Coch, Sanders, & Neville, 2005). Children, who showed a different ERP morphology to the probe stimuli, also showed early attentional modulation within the first 100 ms of processing. This attentional modulation was an amplification of the broad positivity occurring in this time window. In a later study (Sanders, Stevens, Coch, & Neville, 2006), we found that this attention effect was complete by 200 ms in older children age 6–8 years but prolonged through 300 ms in children age 3–5. These data suggest that with sufficient attentional cues, children as young as three years of age are able to attend selectively to an auditory stream and that doing so alters neural activity within 100 ms of processing.

We have employed this paradigm to examine the timing and mechanisms of selective auditory attention in children with specific language impairment (SLI) aged six to eight years and typically developing (TD) control children matched for age, gender, nonverbal IQ, and socioeconomic status (SES) (Stevens, Sanders, & Neville, 2006). As shown in figure 11.3*A,C,* by 100 ms, typically developing children in this study showed an amplification of the sensorineural response to attended as compared to unattended stimuli, just as observed in our larger samples of typically developing children. In contrast, children with SLI showed no evidence of sensorineural modulation with attention, despite behavioral performance indicating that they were performing the task as directed (figure 11.3*B,D*). Moreover, the group differences were specific to signal enhancement (figure 11.4, left).

In a related line of research, we examined the neural mechanisms of selective attention in children from different socioeconomic backgrounds. Previous behavioral studies indicated that children from lower socioeconomic backgrounds experience difficulty with selective attention, particularly in tasks of executive function and tasks that require filtering irrelevant information or suppressing prepotent responses (Farah et al., 2006; Lupien, King, Meaney, & McEwen, 2001; Mezzacappa, 2004; Noble et al., 2005; Noble, McCandliss, & Farah, 2007). Using the same selective auditory attention ERP task described earlier, we observed differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds (Stevens, Lauinger, & Neville, in press). Specifically, children whose mothers had lower levels of education (no college experience) showed reduced effects of selective attention on neural processing compared to children whose mothers had higher levels of education (at least some college) (figure 11.5). These differences were related specifically to a reduced ability to filter irrelevant information (i.e., to suppress the response to ignored sounds) (figure 11.4, right) and could not be accounted for by differences in receptive language skill. Thus the mechanism implicated in attention deficits in children from lower socioeconomic backgrounds (i.e., distractor suppression) was not the same as the mechanism implicated in children with SLI, who showed a deficit in signal enhancement of stimuli in the attended channel (Stevens, Sanders, & Neville, 2006). Similar results have been reported by other research groups (D'Angiulli, Herdman, Stapells, & Hertzman, 2008). Taken together, these studies point to the two sides of the plasticity of early mechanisms of attention, which show both enhancements and vulnerabilities in different populations.

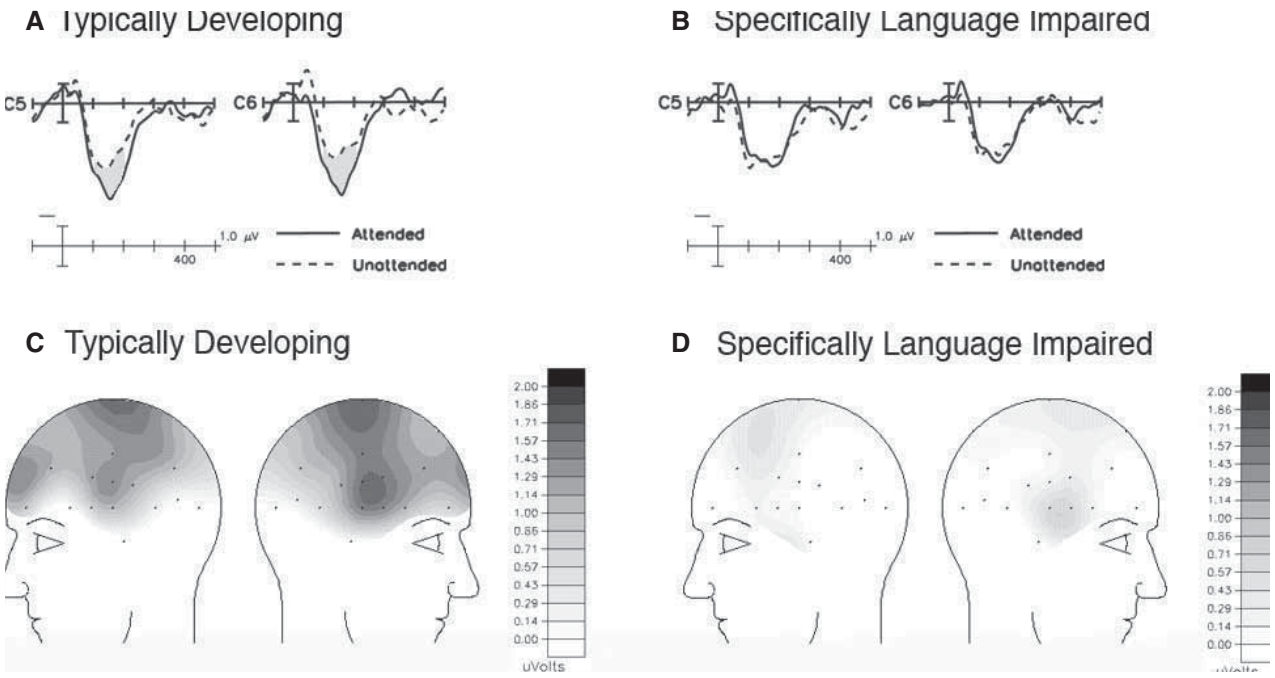


FIGURE 11.3 Grand average event-related potentials (ERPs) for attended and unattended stimuli (*A*) in typically developing children ($P = .001$) and (*B*) in children with specific language impairment ($P > 0.4$). Voltage map of the attention effect (Attended-Unattended) shows (*C*) in typically developing children a large, broadly distributed effect and (*D*) in children with specific language impairment no modulation with attention. (Data from Stevens, Sanders, & Neville, 2006. Image reproduced with permission from *Brain Research*.)

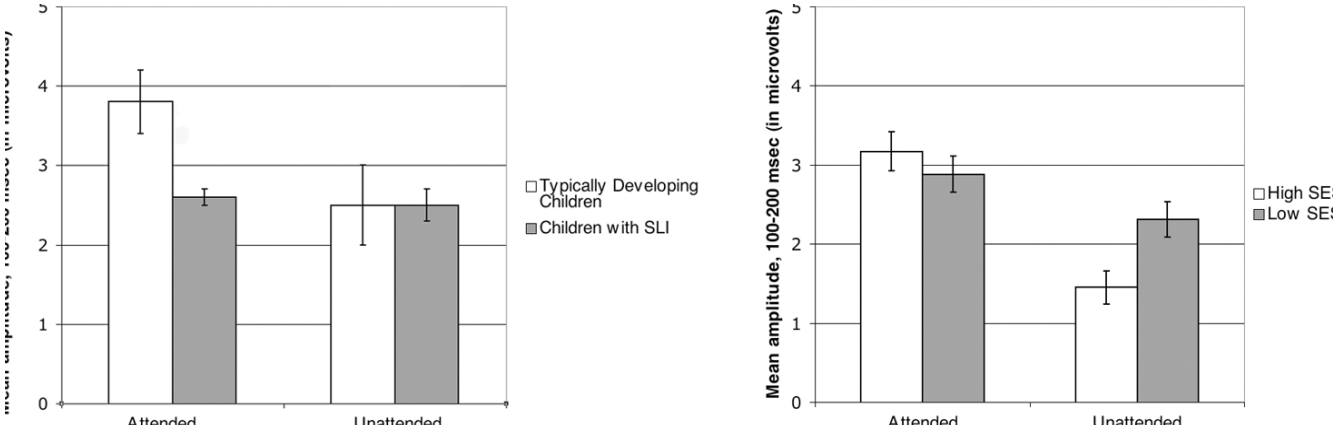


FIGURE 11.4 Mean amplitude of the ERP from 100 to 200 ms of responses to unattended and attended probes. Error bars represent standard error of the mean. Left panel shows data from typically developing children (TD) and children with specific language impairment (SLI). The two groups did not differ in the magnitude of response to unattended stimuli. However, typically developing children showed a larger amplitude response (i.e., better signal enhancement) than children with SLI to attended stimuli. Right panel shows data from children from higher versus lower socioeconomic backgrounds. Children from different socioeconomic backgrounds did not differ in the magnitude of response to attended stimuli. However, children from lower socioeconomic backgrounds showed a larger response (i.e., poorer filtering) to unattended stimuli compared to children from higher socioeconomic backgrounds. (Data from Stevens, Sanders, & Neville, 2006, *Brain Research*, and Stevens, Lauinger, & Neville, in press, *Developmental Science*.)

Several mechanisms might underlie the plasticity of attention. Whereas the research described previously focused on sustained, selective attention, research in cognitive science and cognitive neuroscience has also identified several different subsystems, or components of attention (Coull, Frith, Frackowiak, & Grasby, 1996; Raz & Buhle, 2006; Shipp, 2004). These components of attention depend upon different neural substrates and neurotransmitters (Bush, Luu, & Posner, 2000; Gomes, Molholm, Christodoulou, Ritter, & Cowan, 2000; Posner & Petersen, 1990) and

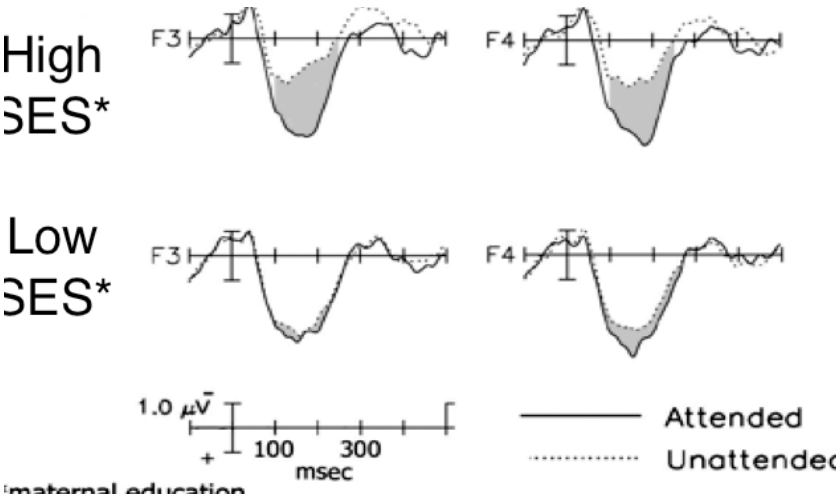


FIGURE 11.5 Grand average evoked potentials for attended and unattended stimuli in children from higher socioeconomic backgrounds (upper panel) and lower socioeconomic backgrounds (lower panel). The effect of attention on sensorineural processing was significantly larger in children from higher socioeconomic backgrounds ($P = .001$). (Data from Stevens, Lauinger, & Neville, in press, *Developmental Science*.)

mature along different timetables (Andersson & Hugdahl, 1987; Doyle, 1973; Geffen & Wale, 1979; Hiscock & Kinsbourne, 1980; Pearson & Lane, 1991; Rueda, Fan, et al., 2004; Rueda, Posner, Rothbart, & Davis-Stober, 2004; Schul, Townsend, & Stiles, 2003). Sustained, selective attention shows a particularly long time course of development. The abilities both to selectively attend to relevant stimuli and to successfully ignore irrelevant stimuli improve progressively with increasing age across childhood (Cherry, 1981; Geffen & Sexton, 1978; Geffen & Wale, 1979; Hiscock & Kinsbourne, 1980; Lane & Pearson, 1982; Maccoby & Konrad, 1966; Sexton & Geffen, 1979; Zukier & Hagen, 1978). Further, there is some evidence that background noise creates greater interference effects for younger children than for adolescents or adults (Elliott, 1979; Ridderinkhof & van der Stelt, 2000). In a review of both behavioral and ERP studies of the development of selective attention, Ridderinkhof and van der Stelt (2000) proposed that the abilities to select among competing stimuli and to preferentially process more relevant information are essentially available in very young children, but that the speed and efficiency of these behaviors and the systems contributing to these abilities improve as children develop. Additionally, since the key sources of selective attention within the parietal and frontal lobes constitute parts of the dorsal pathway, similar chemical and anatomical factors noted in the section on vision may contribute to the plasticity of attention in a similar way. In addition, recent evidence suggests that there are considerable genetic effects on attention (Bell et al., 2008; Fan, Fossella, Sommer, Wu, & Posner, 2003; Posner, Rothbart, & Sheese, 2007; Rueda, Rothbart, McCandliss, Saccamanno, & Posner, 2005) and that these may also be modified by environmental input epigenetically (Bakermans-Kranenberg et al., 2008; Sheese et al., 2007; unpublished observations from our lab).

Interventions

As described in the preceding section, selective attention influences early sensory processing across a number of domains. In our most recent research, we have been investigating the possibility that attention itself might be trainable, and that this training can impact processing in a number of different domains. Indeed, in his seminal work *Principles of Psychology*, William James raised the idea of attention training for children, proposing that this would be "*the* education *par excellence*" (James, 1890, italics in original). While James went on to say that such an education is difficult to define and bring about, attention training has recently been implemented in curricula for preschool and school-age children (Bodrova & Leong, 2007; Chenault, Thomson, Abbot, & Berninger, 2006; Diamond, Barnett, Thomas, & Munro, 2007; Rueda et al., 2005). These programs are associated with improvements in behavioral and neurophysiological indices of attention, as well as in measures of academic outcomes and nonverbal intelligence. Furthermore, one program showed that attention training translated to increased benefits of a subsequent remedial writing intervention for adolescents with dyslexia (Chenault et al., 2006).

Recent proposals suggest that some interventions designed to improve language skills might also target or train selective attention (Gillam, 1999; Gillam, Loeb, & Friel-Patti, 2001; Gillam, Crofford, Gale, & Hoffman, 2001; Hari & Renvall, 2001). We have tested this hypothesis in a series of intervention studies. In this research, we have documented changes in the neural mechanisms of selective attention following

training in typically developing children, as well as children with language impairment or at risk for reading failure (Stevens, Coch, Sanders, & Neville, 2008; Stevens, Harn, et al., in press). In all cases, increases in the effects of attention on sensorineural processing were accompanied by behavioral changes in other domains that were also targeted by the training programs, including language and preliteracy skills. These data suggest that modifications in behavior can arise alongside changes in the early neural mechanisms of attention.

In one study, we examined whether six weeks of high-intensity (100 min/day) training with a computerized intervention program designed to improve language skills would also influence neural mechanisms of selective auditory attention previously shown to be deficient in children with SLI (Stevens, Coch, et al., 2008). Before and after training (or a comparable delay period for a no-treatment control group), children completed standardized language assessments and the ERP measure of selective auditory attention described earlier. Relative to the no-treatment control group, both children with SLI and typically developing children receiving training showed increases in standardized measures of receptive language. In addition, children receiving training showed larger increases in the effects of attention on neural processing following training relative to the control group, and these changes were specific to changes in signal enhancement of attended stimuli (figure 11.6).

In a second study, we examined the neural mechanisms of selective attention in kindergarten children who were either on track in preliteracy skills or at risk for reading failure. They were studied at the beginning of and following the first semester of kindergarten (Stevens, Currin, et al., 2008; Stevens, Harn, et al., in press). The at-risk group also received supplemental instruction with a previously validated reading intervention (Simmons, Kame'enui, Stoolmiller, Coyne, & Harn, 2003; Simmons et al., 2007). Behaviorally, the at-risk group showed improved performance on several preliteracy measures, raising their performance close to the on-track group by the end of the year. At the start of kindergarten, the at-risk group displayed reduced effects of attention on sensorineural processing compared to the on-track group. Following training, this difference between groups disappeared, with the at-risk group showing increased effects of attention on sensorineural processing (figure 11.7).

Functional MRI data from the same kindergarten children further supported the role of attentional changes in successful language or reading interventions (Yamada et al., 2008; Yamada, Stevens, Harn, Chard, & Neville, under review). Hemodynamic responses to visually presented letters or false-font stimuli (presented in separate blocks) were examined. Participants indicated when the same letter or false-font stimulus was repeated in two successive trials (i.e., a 1-back task). Consistent with previous research on reading-related networks in fluent readers, adults recruited a left temporoparietal region during this task (figure 11.8*A*). At the start of kindergarten, on-track children recruited bilateral temporoparietal regions, whereas children at risk for reading

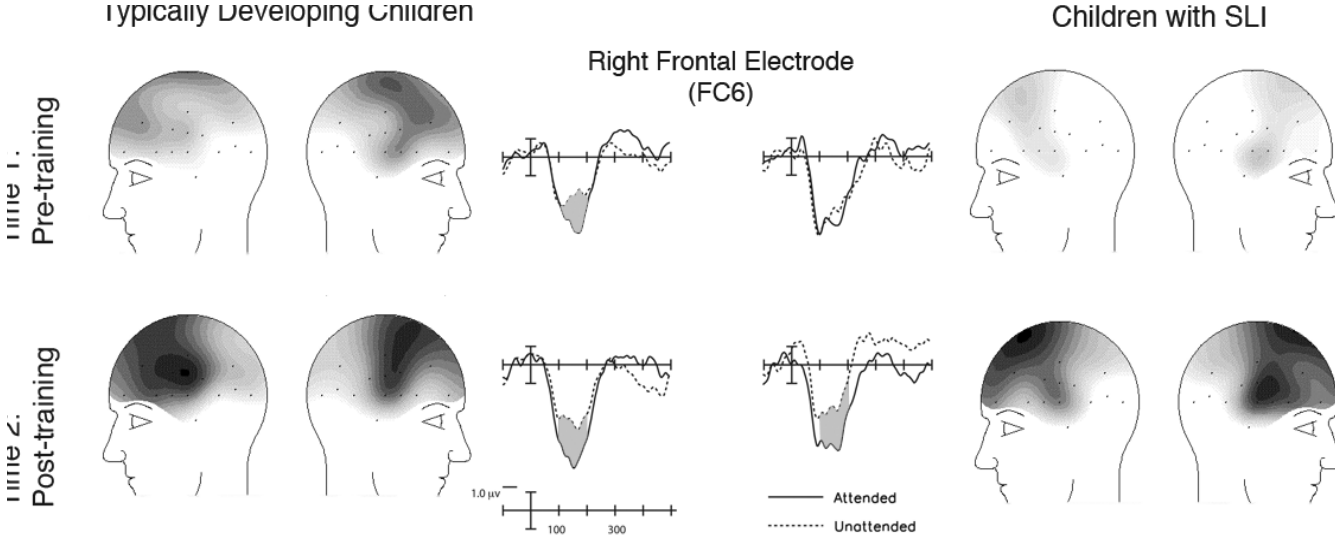


Figure 11.6 ERP responses to attended and ignored auditory stimuli in typically developing (TD) children and children with specific language impairment (SLI) before and after six weeks of daily, 100-minute computerized language training. Grand average evoked potentials for attended and unattended stimuli are collapsed across linguistic and nonlinguistic probes. Voltage maps show magnitude and distribution of the attention effect (attended-unattended) during the 100–200-ms time window. Following training, both children with SLI ($P < .05$) and typically developing children ($P < .1$) showed evidence of increased effects of attention on sensorineural processing. These changes were larger than those made in a no-treatment control group ($P < .01$), who showed no change in the effects of attention on sensorineural processing when retested after a comparable time period ($P = .96$).

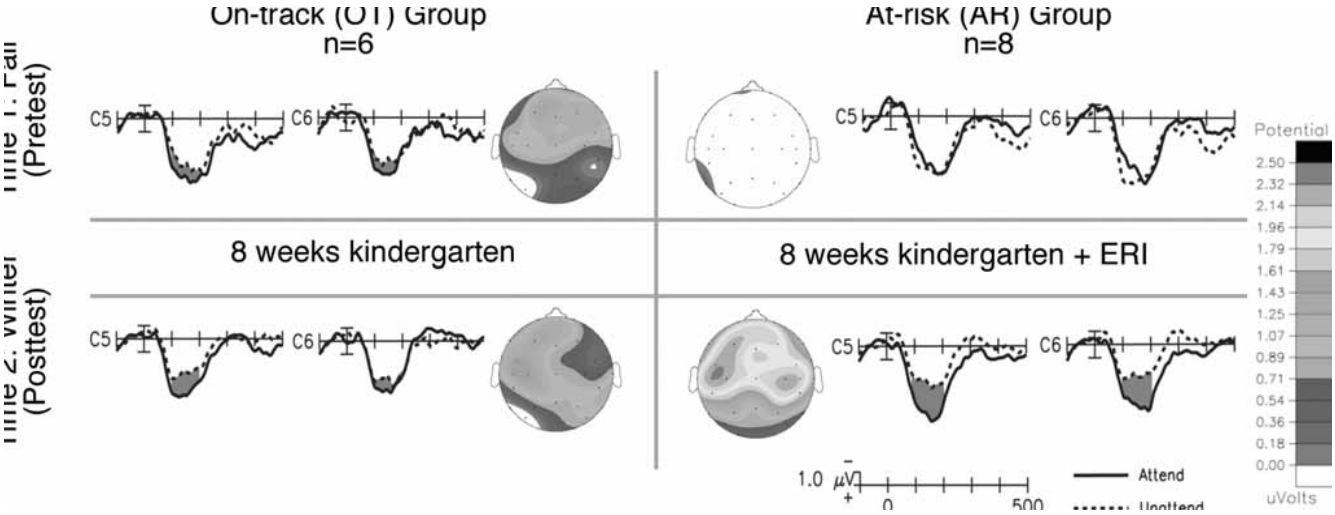


FIGURE 11.7 Grand average ERP waveforms from the selective auditory attention paradigm show the effects of attention on sensorineural processing in kindergarten children of diverse early reading ability across the first semester of kindergarten. Top row shows data from pretest, and bottom row shows data from posttest for five-year-old kindergarten children on track (OT) in early literacy skills or at risk (AR) for reading difficulty. The OT group received eight weeks of kindergarten between pretest and posttest. The AR group received eight weeks of kindergarten with 45 minutes of daily, supplemental instruction with the Early Reading Intervention (ERI). Voltage map indicates the magnitude and distribution of the attention effect (Attended-Unattended). Changes in the effects of attention differed from pretest to posttest in the two groups ($P < .05$), with the OT group showing no change ($P = .92$) and the AR group showing a significant increase in the attention effect ($P < .01$). At pretest, the OT group tended to have a larger attention effect than the AR group ($P = .06$). At posttest, the AR group had a nonsignificantly larger attention effect than the OT group ($P = .17$). (See color plate 12.)

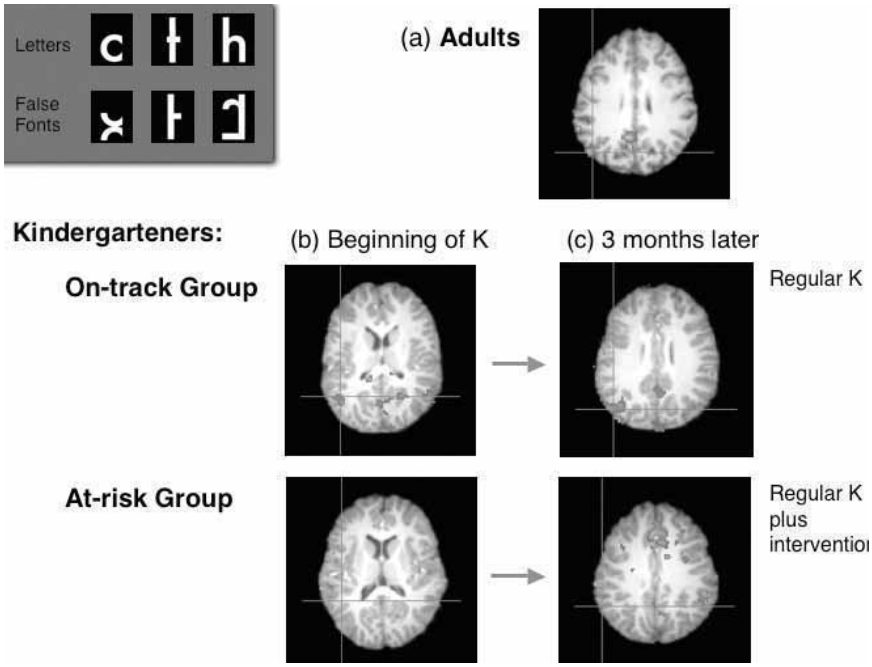


FIGURE 11.8 Functional MRI activations for letter > false font while performing a 1-back task in adults and kindergarten children of diverse reading ability across the first semester of formal reading instruction. (*A*) Adults performing the task displayed activation in classic left temporoparietal regions. (*B*) In contrast, at the beginning of kindergarten, children on track in early literacy skills (upper panel) showed bilateral temporoparietal activation, and children at risk for reading difficulty (lower panel) showed no regions of greater activation. (*C*) Following one semester of kindergarten and, for children in the at-risk group, daily supplemental instruction with the Early Reading Intervention, on-track children showed left-lateralized activation in temporoparietal regions, and at-risk children showed bilateral temporoparietal activation and large activation of frontal regions, including the ACC. The left hemisphere is displayed on the left. In the upper left corner are example stimuli. (See color plate 13.)

failure did not show greater recruitment of any brain regions for letters versus false-font stimuli (figure 11.8*B*). Following 3 months of kindergarten and, for children at risk for reading failure, supplemental reading instruction, both groups showed changes in reading circuits toward more adultlike patterns, though the at-risk group showed a less mature pattern of activation (figure 11.8*C*). Interestingly, following a semester of kindergarten, the at-risk group showed greater activation than the on-track group of supplemental frontal regions, including the anterior cingulate cortex (figure 11.8). This suggested that changes in the neural circuits for reading in response to intervention also involved the recruitment of additional neural resources related to attention.

In a related line of research, we have also begun studies that train parents of children from lower socioeconomic backgrounds. Across eight weekly, small-group sessions, parents learn evidence-based strategies to improve communication with their children, promote children's critical thinking skills, and decrease family stress. We have compared the pre- to posttraining changes in this group of parents and their children to changes in a matched control group randomly assigned not to receive the intervention. To date the parent training appears very promising (Fanning, 2007; Fanning, Paulsen, Sundborg, & Neville, 2008; Fanning, Sohlberg, & Neville, under review). Relative to the control group, parents in the intervention group show larger decreases in self-reported stress related to parenting challenges. When interacting with their children, their language becomes more child directed (e.g., they allow more opportunities for the child to talk and to guide the interaction). In addition, there are large changes in the children themselves. Children whose parents completed the training show large and significant increases in standardized measures of language, nonverbal IQ, memory, and attention compared to children whose parents are randomly assigned not to receive the intervention. We are continuing to assess the parent-training program by looking at the effects of the training on attention and language-related ERPs. We are following children longitudinally to see whether improvements persist and generalize school performance. In addition, we are adapting the parent-training program to include a stronger focus on developing children's attention and self-regulation skills.

Finally, recent studies have linked variability in polymorphisms of genes that influence the production, metabolism, and transport of neurotransmitters important in attention to variability in behavioral, ERP, and fMRI indices of attention (Bell et al., 2008; Fan et al., 2003; Greenwood & Parasuraman, 2003). For example, the 3 repeat allele of the MAOA gene is associated with reductions in language and cognition, including executive attention, and with reductions in our ERP attention effects and activation of ACC on fMRI compared to the 4 repeat allele (Bell et al.; Fan et al.). The long-long allele of the serotonin transporter gene and the 7 repeat allele of the DRD4 gene are associated with increased rates of ADHD and reduction of our ERP attention effects (Bell et al.; Fan et al.; Parasuraman, Greenwood, Kumar, & Fosella, 2005; Rueda et al., 2005; Savitz, Solmes, & Ramesar, 2006). However, recent studies suggest that such genetic effects display plasticity that is dependent on and modified by environmental input including parenting quality, parental interventions, and small group interventions (Bakermans-Kranenberg & Van Ijzendoom, 2006; Bakermans-Kranenberg et al., 2008; Sheese et al., 2007; and our unpublished observations). Thus gene × environment interactions and epigenetic mechanisms similar to those operating in animal studies (Kondo et al., 2008; Suomi, 2003, 2006) likely play a role in determining the different profiles of human neuroplasticity as well.

Conclusions

The research described in this chapter has illustrated the variable degrees and time periods of neuroplasticity in the human brain and likely mechanisms whereby experience influences different subsystems within perceptual and cognitive domains. Additionally, this research has highlighted the bidirectional nature of plasticity—those aspects of neural processing and related cognitive functioning that show the greatest capability for enhancement also display the greatest susceptibility to deficits under different conditions. Researchers are entering an exciting frontier of neuroplasticity research that takes the results of basic research on the profiles and mechanisms of neuroplasticity as a point of departure in the development of training and intervention programs. Our growing understanding of the limits and mechanisms of plasticity contributes to a basic understanding of human brain development and function and can also inform and guide efforts to harness neuroplasticity both to optimize and to protect the malleable and vulnerable aspects of human development.

ACKNOWLEDGMENTS We thank our many collaborators in the research reported here. Supported by grants from NIH NIDCD (DC000481, DC000128) and Department of Education IES (R305B070018).

REFERENCES

Adamson, A., Mills, D., Appelbaum, G., & Neville, H. (1998). *Auditory sentence processing in adults and children: Evidence from ERPs.* Poster presented at the Cognitive Neuroscience Society, San Francisco, CA.

Adamson-Harris, A. M., Mills, D. L., & Neville, H. J. (2000). *Children's processing of grammatical and semantic information within sentences: Evidence from event-related potentials.* Poster presented at the Cognitive Neuroscience Society, San Francisco, CA.

Andersson, B., & Hugdahl, K. (1987). Effects of sex, age, and forced attention on dichotic listening in children: A longitudinal study. *Dev. Neuropsychol.*, *3*(3–4), 191–206.

Armstrong, B., Hillyard, S. A., Neville, H. J., & Mitchell, T. V. (2002). Auditory deprivation affects processing of motion, but not color. *Cogn. Brain Res.*, *14*(3), 422–434.

Atkinson, J. (1991). Review of human visual development: Crowding and dyslexia. In J. Cronly-Dillon & J. Stein (Eds.), *Vision and visual dysfunction* (Vol. 13, pp. 44–57). London: Nature Publishing Group.

Atkinson, J. (1992). Early visual development: Differential functioning of parvocellular and magnocellular pathways. *Eye*, *6*, 129–135.

Atkinson, J., King, J., Braddick, O., Nokes, L., Anker, S., & Braddick, F. (1997). A specific deficit of dorsal stream function in Williams' syndrome. *Neuroreport*, *8*(8), 1919–1922.

Baizer, J. S., Ungerleider, L. G., & Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *J. Neurosci.*, *11*, 168–190.

Bakermans-Kranenberg, M., & Van Ijzendoom, M. H. (2006). Gene-environment interaction of the dopamine D4 receptor (DRD4) and observed maternal insensitivity predicting externalizing behavior in preschoolers. *Dev. Psychobiol.*, *48*, 406–409.

Bakermans-Kranenberg, M., Van Ijzendoom, M. H., Pijlman, F. T. A., Mesman, J., & Femmie, J. (2008). Experimental evidence for differential susceptibility: Dopamine D4 receptor polymorphism (DRD4 VNTR) moderates intervention effects on toddlers' externalizing behavior in a randomized controlled trial. *Dev. Psychol.*, *44*, 293–300.

Bavelier, D., Brozinsky, C., Tomann, A., Mitchell, T., Neville, H., & Liu, G. (2001). Impact of early deafness and early exposure to sign language on the cerebral organization for motion processing. *J. Neurosci.*, *21*(22), 8931–8942.

Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: Where and how? *Nat. Rev. Neurosci.*, *3*, 443–452.

Bavelier, D., Tomann, A., Hutton, C., Mitchell, T., Liu, G., Corina, D., et al. (2000). Visual attention to the periphery is enhanced in congenitally deaf individuals. *J. Neurosci.*, *20*(17), 1–6.

Bell, T., Batterink, L., Currin, L., Pakulak, E., Stevens, C., & Neville, H. (2008). *Genetic influences on selective auditory attention as indexed by ERPs*. Poster presented at the Cognitive Neuroscience Society, San Francisco.

Bishop, D. (2003). Genetic and environmental risks for specific language impairment in children. *Int. J. Pediatr. Otorhinolaryngol.*, *67*, S143–157.

Bishop, D., Bishop, S. J., Bright, P., James, C., Delaney, T., & Tallal, P. (1999). Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. *J. Speech Lang. Hear. Res.*, *42*, 155–168.

Bishop, D., & McArthur, G. M. (2004). Immature cortical responses to auditory stimuli in specific language impairment: Evidence from ERPs to rapid tone sequences. *Dev. Sci.*, *7*, F11–18.

Bodrova, E., & Leong, D. (2007). *Tools of the mind: The Vygotskian approach to early childhood education* (2nd ed.). Upper Saddle River, NJ: Pearson Education.

Burton, H., Synder, A., Conturo, T., Akbudak, E., Ollinger, J., & Raichle, M. (2002). *A fMRI study of verb generation to auditory nouns in early and late blind*. Poster presented at the Society for Neuroscience, Orlando, FL.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.*, *4*(6), 215–222.

Capek, C. (2004). *The cortical organization of spoken and signed sentence processing in adults*. Unpublished doctoral dissertation, University of Oregon, Eugene.

Capek, C., Bavelier, D., Corina, D., Newman, A. J., Jezzard, P., & Neville, H. J. (2004). The cortical organization for audio-visual sentence comprehension: An fMRI study at 4 Tesla. *Cogn. Brain Res.*, *20*(2), 111–119.

Capek, C., Corina, D., Grossi, G., McBurney, S. L., Mitchell, T. V., Neville, H. J., et al. (under review). Semantic and syntactic processing in American Sign Language: Electrophysiological evidence.

Capek, C., Corina, D., Grossi, G., McBurney, S. L., Neville, H. J., Newman, A. J., et al. (in preparation). American Sign Language sentence processing: ERP evidence from adults with different ages of acquisition.

Chalupa, L. M., & Dreher, B. (1991). High precision systems require high precision "blueprints": A new view regarding the formation of connections in the mammalian visual system. *J. Cogn. Neurosci.*, *3*(3), 209–219.

Chenault, B., Thomson, J., Abbott, R. D., & Berninger, V. W. (2006). Effects of prior attention training on child dyslexics' response to composition instruction. *Dev. Neuropsychol.*, *29*(1), 243–260.

Cherry, R. (1981). Development of selective auditory attention skills in children. *Percept. Mot. Skills*, *52*, 379–385.

Chugani, H. T., Phelps, M. E., & Mazziotta, J. C. (1987). Positron emission tomography study of human brain functional development. *Ann. Neurol.*, *22*, 487–497.

Coch, D., Sanders, L. D., & Neville, H. J. (2005). An event-related potential study of selective auditory attention in children and adults. *J. Cogn. Neurosci.*, *17*(4), 605–622.

Coch, D., Skendzel, W., Grossi, G., & Neville, H. (2005). Motion and color processing in school-age children and adults: An ERP study. *Dev. Sci.*, *8*(4), 372–386.

Cohen, L. G., Weeks, R. A., Celnik, P., & Hallett, M. (1999). Role of the occipital cortex during Braille reading in subjects with blindness acquired late in life. *J. Neurosci.*, *45*, 451–460.

Cooper, H., Herbin, M., & Nevo, E. (1993). Visual system of a naturally microphthalmic mammal: The blind mole rat, *Spalax ehrenbergi*. *J. Comp. Neurol.*, *328*, 313–350.

Corbetta, M., Miezin, F., Dobmeyer, S., Shulman, G., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, *248*, 1556–1559.

Cornelissen, P., Richardson, A., Mason, A., Fowler, S., & Stein, J. (1995). Contrast sensitivity and coherent motion detection measured at photopic luminance levels in dyslexics and controls. *Vis. Res.*, *35*(10), 1483–1494.

Coull, J. T., Frith, C. D., Frackowiak, R. S., & Grasby, P. M. (1996). A fronto-parietal network for rapid visual information processing: A PET study of sustained attention and working memory. *Neuropsychologia*, *34*, 1085–1095.

D'Angiulli, A., Herdman, A., Stapells, D., & Hertzman, C. (2008). Children's event-related potentials of auditory selective attention vary with their socioeconomic status. *Neuropsychology*, *22*, 293–300.

Demb, J. B., Boynton, G. M., Best, M., & Heeger, D. J. (1998). Psychophysical evidence for a magnocellular pathway deficit in dyslexia. *Vis. Res.*, *38*, 1555–1559.

DeYoe, E. A., Hockfield, S., Garren, H., & Van Essen, D. C. (1990). Antibody labeling of functional subdivisions in visual cortex: Cat-301 immunoreactivity in striate and extrastriate cortex of the macaque monkey. *Visual Neurosci.*, *5*, 67–81.

Diamond, A., Barnett, W., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*, 1387–1388.

Doron, N., & Wollberg, Z. (1994). Cross-modal neuroplasticity in the blind mole rat *Spalax ehrenberg*: A WGA-HRP tracing study. *NeuroReport*, *5*, 2697–2701.

Dow, M., Scott, G., Stevens, C., & Neville, H. (2006). *Functional magnetic resonance imaging (fMRI) evidence for distributed visual neuroplasticity in congenitally deaf humans*. Poster presented at the Society for Neuroscience, Atlanta, GA.

Doyle, A. B. (1973). Listening to distraction: A developmental study of selective attention. *J. Exp. Child Psychol.*, *15*, 100–115.

Eckert, M. A., Kamdar, N. V., Chang, C. E., Beckmann, C. F., Greicius, M. D., & Menon, V. (2008). A cross-model system linking primary auditory and visual cortices: Evidence from intrinsic fMRI connectivity analysis. *Hum. Brain Mapping*, *29*, 848–857.

Eden, G. F., VanMeter, J. W., Rumsey, J. M., Maisog, J. M., Woods, R. P., & Zeffiro, T. A. (1996). Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. *Nature*, *382*(6586), 66–69.

Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *J. Acoust. Soc. Am.*, *66*(3), 651–653.

Everatt, J., Bradshaw, M. F., & Hibbard, P. B. (1999). Visual processing and dyslexia. *Perception*, *28*, 243–254.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J. Neurosci.*, *22*(13), 5749–5759.

Fan, J., Fossella, J., Sommer, T., Wu, Y., & Posner, M. I. (2003). Mapping the genetic variation of executive attention onto brain activity. *Proc. Natl. Acad. Sci. USA*, *100*(12), 7406–7411.

Fanning, J. (2007). *Parent training for caregivers of typically developing, economically disadvantaged preschoolers: An initial study in enhancing language development, avoiding behavior problems, and regulating family stress*. Unpublished dissertation, University of Oregon, Eugene.

Fanning, J., Paulsen, D., Sundborg, S., & Neville, H. (2008). *The effects of parent training: Enhancing children's neurocognitive function*. Poster presented at the Cognitive Neuroscience Society, San Francisco.

Fanning, J., Sohlberg, M. M., & Neville, H. (under review). Parent training for caregivers of typically developing, economically disadvantaged preschoolers: Enhancing language development, avoiding behavior problems, and regulating family stress.

Farah, M., Shera, D., Savage, J., Betancourt, L., Giannetta, J., Brodsky, N., et al. (2006). Childhood poverty: Specific associations with neurocognitive development. *Brain Res.*, *1110*, 166–174.

Fieger, A., Röder, B., Teder-Sälejärvi, W., Hillyard, S. A., & Neville, H. J. (2006). Auditory spatial tuning in late onset blind humans. *J. Cogn. Neurosci.*, *18*(2), 149–157.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.*, *6*(2), 78–84.

Galaburda, A., & Livingstone, M. (1993). Evidence for a magnocellular defect in developmental dyslexia. In P. Tallal, A. M. Galaburda, R. R. Llinas, & C. von Euler (Eds.), *Temporal information processing in the nervous system* (pp. 70–82). New York: New York Academy of Sciences.

Galaburda, A., Menard, M., & Rosen, G. (1994). Evidence for aberrant auditory anatomy in developmental dyslexia. *Proc. Natl. Acad. Sci. USA*, *91*, 8010–8013.

Garel, S., Huffman, K. J., & Rubenstein, J. L. R. (2003). Molecular regionalization of the neocortex is disrupted in Fgf8 hypomorphic mutants. *Development*, *130*, 1903–1914.

Geffen, G., & Sexton, M. A. (1978). The development of auditory strategies of attention. *Dev. Psychol.*, *14*(1), 11–17.

Geffen, G., & Wale, J. (1979). Development of selective listening and hemispheric asymmetry. *Dev. Psychol.*, *15*(2), 138–146.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nat. Neurosci.*, *2*(10), 861–863.

Gillam, R. (1999). Computer-assisted language intervention using Fast ForWord: Theoretical and empirical considerations for clinical decision-making. *Lang. Speech Hear. Serv. Schools*, *30*, 363–370.

Gillam, R., Crofford, J., Gale, M., & Hoffman, L. (2001). Language change following computer-assisted language instruction with Fast ForWord or Laureate Learning Systems software. *Am. J. Speech-Lang. Pathol.*, *10*, 231–247.

Gillam, R., Loeb, D. F., & Friel-Patti, S. (2001). Looking back: A summary of five exploratory studies of Fast ForWord. *Am. J. Speech-Lang. Pathol.*, *10*, 269–273.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. USA*, *101*, 8174–8179.

Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., & Cowan, N. (2000). The development of auditory attention in children. *Front. Biosci.*, *5*, 108–120.

Greenwood, P. M., & Parasuraman, R. (2003). Normal genetic variation, cognition, and aging. *Behav. Cogn. Neurosci. Rev.*, *2*, 278–306.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *J. Cogn. Neurosci.*, *16*(7), 1302–1318.

Hansen, P. C., Stein, J. F., Orde, S. R., Winter, J. L., & Talcott, J. B. (2001). Are dyslexics' visual deficits limited to measures of dorsal stream function? *NeuroReport*, *12*(7), 1527–1530.

Hari, R., & Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends Cogn. Sci.*, *5*(12), 525–532.

Heil, P., Bronchti, G., Wollberg, Z., & Scheich, H. (1991). Invasion of visual cortex by the auditory system in the naturally blind mole rat. *NeuroReport*, *2*(12), 735–738.

Hickey, T. L. (1981). The developing visual system. *Trends Neurosci.*, *2*, 41–44.

Hillyard, S., Di Russo, F., & Martinez, A. (2003). Imaging of visual attention. In N. Kanwisher & J. Duncan (Eds.), *Attention and performance XX: Functional neuroimaging of visual cognition, attention, and performance*, Oxford, UK: Oxford University Press.

Hillyard, S., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signals of selective attention in the human brain. *Science*, *182*(4108), 177–179.

Hiscock, M., & Kinsbourne, M. (1980). Asymmetries of selective listening and attention switching in children. *Dev. Psychol.*, *16*(1), 70–82.

Hockfield, S. (1983). A surface antigen expressed by a subset of neurons in the vertebrate central nervous system. *Proc. Natl. Acad. Sci. USA*, *80*(18), 5758–5761.

Hollants-Gilhuijs, M. A. M., Ruijter, J. M., & Spekreijse, H. (1998a). Visual half-field development in children: Detection of colour-contrast-defined forms. *Vis. Res.*, *38*(5), 645–649.

Hollants-Gilhuijs, M. A. M., Ruijter, J. M., & Spekreijse, H. (1998b). Visual half-field development in children: Detection of motion-defined forms. *Vis. Res.*, *38*(5), 651–657.

Hunt, D. L., King, B., Kahn, D. M., Yamoah, E. N., Shull, G. E., & Krubitzer, L. (2005). Aberrant retinal projections in congenitally deaf mice: How are phenotypic characteristics specified in development and evolution? *Anat. Rec. A Discov. Mol. Cell Evol. Biol.*, *287*(1), 1051–1066.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.*, *387*, 167–178.

Hyvarinen, L., & Linnankoski, I. (1981). Modification of parietal association cortex and functional blindness after binoocular deprivation in young monkeys. *Exp. Brain Res.*, *42*, 1–8.

James, W. (1890). *Principles of psychology*. New York: Henry Holt.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning; The influence of maturational state on the acquisition of English as a second language. *Cogn. Psych.*, *21*, 60–99.

Kondo, M., Gray, L., Pelka, G., Christodoulou, J., Tam, P., & Hannan, A. (2008). Environmental enrichment ameliorates a motor coordination deficit in a mouse model of Rett syndrome—Mecp2 gene dosage effects and BDNF expression. *Eur. J. Neurosci.*, *27*, 3342–3350.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–204.

Lane, D., & Pearson, D. (1982). The development of selective attention. *Merrill-Palmer Q.*, *28*(3), 317–337.

Lange, K., & Röder, B. (2005). Orienting attention to points in time improves stimulus processing both within and across modalities. *J. Cogn. Neurosci.*, *18*(5), 715–729.

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. *Psychophysiology*, *40*, 806–817.

Lipina, S., Martelli, M., Vuelta, B., & Colombo, J. (2005). Performance on the A-not-B task of Argentinian infants from unsatisfied and satisfied basic needs homes. *Interamerican J. Psychol.*, *39*, 49–60.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement and depth: Anatomy, physiology, and perception. *Science*, *240*, 740–749.

Lovegrove, W., Martin, F., & Slaghuis, W. (1986). A theoretical and experimental case for a visual deficit in specific reading disability. *Cogn. Neuropsychol.*, *3*, 225–267.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cogn. Sci.*, *4*(11), 432–440.

Lupien, S. J., King, S., Meaney, M. J., & McEwen, B. S. (2001). Can poverty get under your skin? Basal cortisol levels and cognitive function in children from low and high socioeconomic status. *Dev. Psychopathol.*, *13*, 653–676.

Maccoby, E., & Konrad, K. (1966). Age trends in selective listening. *J. Exp. Child Psychol.*, *3*, 113–122.

Mangun, G., & Hillyard, S. (1990). Electrophysiological studies of visual selective attention in humans. In A. Scheibel & A. Wechsler (Eds.), *Neurobiology* of *higher cognitive function* (pp. 271–295). New York: Guilford.

Mayberry, R. (1993). First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *J. Speech Hear. Res.*, *36*(6), 1258–1270.

Mayberry, R. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain Lang.*, *87*(3), 369–384.

Mayberry, R., & Eichen, E. (1991). The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *J. Mem. Lang.*, *30*, 486–512.

Mayberry, R., Lock, E., & Kazmi, H. (2002). Linguistic ability and early language exposure. *Nature*, *417*, 38.

Merigan, W. H. (1989). Chromatic and achromatic vision of macaques: Role of the P pathway. *J. Neurosci.*, *9*(3), 776–783.

Merigan, W. H., & Maunsell, J. (1990). Macaque vision after magnocellular lateral geniculate lesions. *Visual Neurosci.*, *5*, 347–352.

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in epidemiological sample of young, urban children. *Child Dev.*, *75*(5), 1373–1386.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *J. Cogn. Neurosci.*, *5*(3), 317–334.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Dev. Neuropsychol.*, *13*(3), 397–445.

Mitchell, T. V., & Neville, H. J. (2004). Asynchronies in the development of electrophysiological responses to motion and color. *J. Cogn. Neurosci.*, *16*(8), 1–12.

Movshon, J. A., & Blakemore, C. (1974). Functional reinnervation in kitten visual cortex. *Nature*, *251*(5474), 504–505.

Neville, H. J. (1998). Human brain development. In M. Posner & L. Ungerleider (Eds.), *Fundamental Neuroscience* (pp. 1313–1338). New York: Academic Press.

Neville, H. J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A., et al. (1998). Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. *Proc. Natl. Acad. Sci. USA*, *95*(3), 922–929.

Neville, H. J., Coffey, S. A., Holcomb, P. J., & Tallal, P. (1993). The neurobiology of sensory and language processing in language-impaired children. *J. Cogn. Neurosci.*, *5*(2), 235–253.

Neville, H. J., & Lawson, D. (1987a). Attention to central and peripheral visual space in a movement detection task: An event-related potential and behavioral study. I. Normal hearing adults. *Brain Res.*, *405*, 253–267.

Neville, H. J., & Lawson, D. (1987b). Attention to central and peripheral visual space in a movement detection task: An event-related potential and behavioral study. II. Congenitally deaf adults. *Brain Res.*, *405*, 268–283.

Neville, H. J., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *J. Cogn. Neurosci.*, *3*, 155–170.

Neville, H. J., Schmidt, A., & Kutas, M. (1983). Altered visual-evoked potentials in congenitally deaf adults. *Brain Res.*, *266*, 127–132.

Newman, A. J., Bavelier, D., Corina, D., Jezzard, P., & Neville, H. J. (2002). A critical period for right hemisphere recruitment in American Sign Language processing. *Nature Neurosci.*, *5*(1), 76–80.

Newman, A. J., Ullman, M. T., Pancheva, R., Waligura, D., & Neville, H. (2007). An ERP study of regular and irregular English past tense inflection. *NeuroImage*, *34*(1), 435–445.

Noble, K., McCandliss, B., & Farah, M. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Dev. Sci.*, *10*, 464–480.

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Dev. Sci.*, *8*(1), 74–87.

Packer, O., Hendrickson, A., & Curcio, A. (1990). Developmental redistribution of photoreceptors across the *Macaca nemestrina* (pigtail macaque) retina. *J. Comp. Neurol.*, *298*, 472–493.

Pakulak, E., Hyde, S., Jackobs, Z., & Neville, H. (2007). *Individual differences in syntactic processing as revealed by ERPs and fMRI.* Poster presented at the Cognitive Neuroscience Society, New York.

Pakulak, E., & Neville, H. (under review). Proficiency differences in syntactic processing of native speakers indexed by event-related potentials.

Parasuraman, R., Greenwood, P. M., Kumar, R., & Fosella, J. (2005). Beyond heritability: Neurotransmitter genes differentially modulate visuospatial attention and working memory. *Psychol. Sci.*, *16*, 200–207.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annu. Rev. Neurosci.*, *28*, 377–401.

Pearson, D. A., & Lane, D. M. (1991). Auditory attention switching: A developmental study. *J. Exp. Child Psychol.*, *51*, 320–334.

Posner, M., & Petersen, S. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.*, *13*, 25–42.

Posner, M., Rothbart, M. K., & Sheese, B. E. (2007). Attention genes. *Dev. Sci.*, *10*, 24–29.

Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiol. Neuro-Otol.*, *3*, 86–103.

Raz, A., & Buhle, J. (2006). Typologies of attention networks. *Nat. Rev. Neurosci.*, *7*, 367–379.

Recanzone, G., Jenkins, W., Hradek, G., & Merzenich, M. (1992). Progressive improvement in discriminative abilities in adult owl monkeys performing a tactile frequency discrimination task. *J. Neurophysiol.*, *67*, 1015–1030.

Recanzone, G., Schreiner, C., & Merzenich, M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.*, *12*, 87–103.

Ridderinkhof, K. R., & van der Stelt, O. (2000). Attention and selection in the growing child: Views derived from developmental psychophysiology. *Biol. Psychol.*, *54*, 55–106.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *Int. J. Psychophysiol.*, *50*, 19–26.

Röder, B., Rösler, F., Hennighausen, E., & Näcker, F. (1996). Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects. *Cogn. Brain Res.*, *4*, 77–93.

Röder, B., Stock, O., Bien, S., Neville, H., & Rösler, F. (2002). Speech processing activates visual cortex in congenitally blind humans. *Eur. J. Neurosci.*, *16*, 930–936.

Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature*, *400*(6740), 162–166.

Rueda, M., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., et al. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*, 1029–1040.

Rueda, M., Posner, M. I., Rothbart, M. K., & Davis-Stober, C. P. (2004). Development of the time course for processing conflict: An event-related potentials study with 4 year olds and adults. *BMC Neurosci*, *5*, 1–13.

Rueda, M., Rothbart, M., McCandliss, B., Saccamanno, L., & Posner, M. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proc. Natl. Acad. Sci., USA*, *102*, 14931–14936.

Sabourin, L., Pakulak, E., Paulsen, D., Fanning, J. L., & Neville, H. (2007). *The effects of age, language proficiency and SES on ERP indices of syntactic processing in children.* Poster presented at the Cognitive Neuroscience Society, New York.

Sanders, L., & Astheimer, L. (2008). Temporally selective attention modulates early perceptual processing: Event-related potential evidence. *Percept. Psychophys*, *7*, 732–742.

Sanders, L., & Neville, H. J. (2003a). An ERP study of continuous speech processing. I. Segmentation, semantics, and syntax in native English speakers. *Cogn. Brain Res.*, *15*(3), 228–240.

Sanders, L., & Neville, H. (2003b). An ERP study of continuous speech processing. II. Segmentation, semantics, and syntax in non-native speakers. *Cogn. Brain Res.*, *15*(3), 214–227.

Sanders, L., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.*, *5*(7), 700–703.

Sanders, L., Stevens, C., Coch, D., & Neville, H. J. (2006). Selective auditory attention in 3- to 5-year-old children: An event-related potential study. *Neuropsychologia*, *44*, 2126–2138.

Savitz, J., Solms, M., & Ramesar, R. (2006). The molecular genetics of cognition: Dopamine, COMT and BDNF. *Genes Brain Behav.*, *5*, 311–328.

Schiller, P. H., & Malpeli, J. G. (1978). Functional specificity of lateral geniculate nucleus laminae of the rhesus monkey. *J. Neurophysiol.*, *41*(3), 788–797.

Schul, R., Townsend, J., & Stiles, J. (2003). The development of attentional orienting during the school-age years. *Dev. Sci.*, *6*(3), 262–272.

Scott, G., Dow, M. W., & Neville, H. J. (2003). *Human retinotopic mapping of the far periphery.* Poster presented at the Society for Neuroscience, New Orleans, LA.

Scott, G., Dow, M., Stevens, C., & Neville, H. (under review). A network of crossmodal plasticity in the congenitally deaf. *Cerebral Cortex.*

Sedato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Delber, M.-P., Dold, G., et al. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, *380*(11), 526–528.

Sexton, M. A., & Geffen, G. (1979). Development of three strategies of attention in dichotic monitoring. *Dev. Psychol.*, *15*(3), 299–310.

Sheese, B. E., Voelker, P. M., Rothbart, M. K., & Posner, M. I. (2007). Parenting quality interacts with genetic variation in dopamine receptor D4 to influence temperament in early childhood. *Dev. Psychopathol.*, *19*, 1039–1046.

Shipp, S. (2004). The brain circuitry of attention. *Trends Cogn. Sci.*, *8*(5), 223–230.

Simmons, D., Kame'enui, E., Harn, B., Coyne, M., Stoolmiller, M., Edwards, L., et al. (2007). Attributes of effective and economic kindergarten reading intervention: An examination of instructional time and design specificity. *J. Learn. Disabil. 40*, 331–347.

Simmons, D., Kame'enui, E. J., Stoolmiller, M., Coyne, M. D., & Harn, B. (2003). Accelerating growth and maintaining proficiency: A two-year intervention study of kindergarten and first-grade children at risk for reading difficulties. In B. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 197–228). Timonium, MD: York Press.

Sperling, A., Lu, Z.-l., Manis, F. R., & Seidenberg, M. S. (2003). Selective magnocellular deficits in dyslexia: A "phantom contour" study. *Neuropsychologia*, *41*(10), 1422–1429.

Sperling, A., Lu, Z., Manis, F. R., & Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nat. Neurosci.*, *8*, 862–863.

Stevens, C., Fanning, J., Coch, D., Sanders, L., & Neville, H. (2008). Neural mechanisms of selective auditory attention are enhanced by computerized training: Electrophysiological evidence from language-impaired and typically developing children. *Brain Res. 1205*, 55–69.

Stevens, C., Currin, J., Paulsen, D., Harn, B., Chard, D., Larsen, D., et al. (2008). *Kindergarten children at-risk for reading failure: Electrophysiological measures of selective auditory attention before and after the early reading intervention.* Poster presented at the Cognitive Neuroscience Society, San Francisco.

Stevens, C., Harn, H., Chard, D., Currin, J., Parisi, D., & Neville, H. (in press). Examining the role of attention and instruction in at-risk kindergarteners: Electrophysiological measures of selective auditory attention before and after an early literacy intervention. *J. Learn. Disabil.*

Stevens, C., Lauinger, B., & Neville, H. (in press). Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds: An event-related brain potential study. *Dev. Sci.*

Stevens, C., & Neville, H. (2006). Neuroplasticity as a double-edged sword: Deaf enhancements and dyslexic deficits in motion processing. *J. Cogn. Neurosci.*, *18*(5), 701–704.

Stevens, C., Sanders, L., Andersson, A., & Neville, H. (2006). *Vulnerability and plasticity of selective auditory attention in children: Evidence from language-impaired and second-language learners.* Poster presented at the Cognitive Neuroscience Society, San Francisco.

Stevens, C., Sanders, L., & Neville, H. (2006). Neurophysiological evidence for selective auditory attention deficits in children with specific language impairment. *Brain Res.*, *1111*, 143–152.

Suomi, S. (2003). Gene-environment interactions and the neurobiology of social conflict. *Ann. NY Acad. Sci.*, *1008*, 132–139.

Suomi, S. (2004). How gene-environment interactions can influence emotional development in rhesus monkeys. In C. Barcia-Coll, E. L. Bearer, & R. M. Lerner (Eds.), *Nature and nurture: The complex interplay of genetic and environmental influences on human behavior and development* (pp. 35–51). Mahwah, NJ: Lawrence Erlbaum.

Suomi, S. (2006). Risk, resilience, and gene × environment interactions in rhesus monkeys. *Ann. NY Acad. Sci.*, *1994*, 52–62.

Talcott, J. B., Hansen, P. C., Assoku, E. L., & Stein, J. F. (2000). Visual motion sensitivity in dyslexia: Evidence for temporal and energy integration deficits. *Neuropsychologia*, *38*(7), 935–943.

Tallal, P. (1975). Perceptual and linguistic factors in the language impairment of developmental dysphasics: An experimental investigation with the Token Test. *Cortex*, *11*, 196–205.

Tallal, P. (1976). Rapid auditory processing in normal and disordered language development. *J. Speech Hear. Res.*, *19*, 561–571.

Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia*, *12*, 83–93.

Ungerleider, L., & Haxby, J. V. (1994). "What" and "where" in the human brain. *Curr. Opin. Neurobiol.*, *4*, 157–165.

Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Webb, S. J., Monk, C. S., & Nelson, C. A. (2001). Mechanisms of postnatal neurobiological development: Implications for human development. *Dev. Neuropsychol.*, *19*(2), 147–171.

Weber-Fox, C. M., & Neville, H. J. (1996). Neural systems for language processing: Effects of delays in second language exposure. *Brain Cogn.*, *30*, 264–265.

Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L. G., et al. (2000). A positron emission tomographic study of auditory localization in the congenitally blind. *J. Neurosci.*, *20*(7), 2664–2672.

Woods, D. (1990). The physiological basis of selective attention: Implications of event-related potential studies. In J. Rohrbaugh, R. Parasuraman, & R. Johnson (Eds.), *Event-related brain potentials: Issues and interdisciplinary vantages*. New York: Oxford Press.

Yamada, Y., Stevens, C., & Neville, H. (under review). Developmental changes in cortical activations during visual letter processing in kindergarteners: An fMRI study.

Yamada, Y., Stevens, C., Sabourin, L., Klein, S. A., Dow, M., Paulsen, D., et al. (2008). *Changes in cortical activations during visual letter processing across the kindergarten year: A longitudinal fMRI study.* Poster presented at the Cognitive Neuroscience Society, San Francisco.

Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. J. (1991). A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.*, *11*(3), 641–649.

Ziegler, J. C., Pech-Georgel, C., George, F., Alanio, F. X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proc. Natl. Acad. Sci. USA*, *102*(39), 14110–14115.

Zukier, H., & Hagen, J. W. (1978). The development of selective attention under distracting conditions. *Child Dev.*, *49*, 870–873.

III

ATTENTION

Chapter 12 TREISMAN 189

13 KASTNER, McMAINS, AND BECK 205

14 CORBETTA, SYLVESTER, AND SHULMAN 219

15 HOPF, HEINZE, SCHOENFELD, AND HILLYARD 235

16 MANGUN, SARON, AND WALSH 251

17 KARNATH 259

18 ROBERTSON 269

19 MAUNSELL 281

20 WOMELSDORF AND FRIES 289

Introduction

STEVEN J. LUCK AND GEORGE R. MANGUN

THE TERM "ATTENTION" has a broad set of meanings in everyday language, but they are all related in some way to the concept of focusing mental processes. Focusing on one task to the exclusion of others is often called "concentration" in everyday language. This kind of attention is typically studied under the heading of executive control. A good example is the Stroop task, in which the challenge is to perform a relatively unpracticed task (saying the name of a color) while suppressing interference from a highly practiced task (reading a word). This variety of attention is discussed in part IX, on higher cognitive functions (chapters 69–76).

Focusing on one source of sensory inputs to the exclusion of others is termed selective attention or, as a shorthand, simply "selection." A good example is the visual search task, in which the challenge is to find a target object while suppressing interference from distractor objects that may be perceptually similar to the target. This is the variety of attention that is the focus of this section. Although selective attention plays a key role in many sensory modalities, the largest body of work has been in the visual system because our detailed knowledge of the anatomy and physiology of this sensory system provides solid footing for the study of the more ephemeral topic of attention. Visual attention is therefore the main focus of the chapters in this section.

Perhaps the most fundamental distinction in the study of attention is between the control of attention and the implementation of attention. Attentional control processes are responsible for taking general task instructions (e.g., find a grapefruit on the left side of this photograph) and converting these into a bias toward an appropriate set of features (e.g., colors, shapes, locations, etc.). Once a set of potentially relevant features, objects, or locations has been found, attentional implementation processes are responsible for ensuring that these features, objects, or locations receive preferential

processing. In the context of the classic spotlight metaphor, control processes are responsible for directing the attentional beam, and attention is implemented by means of the illumination of objects by the beam.

As discussed in Treisman's chapter on the basic behavioral phenomena of attention, early research on attention focused on the implementation of attention (chapter 12). This research initially asked whether attention operates at an early, perceptual stage or a late, postperceptual stage, but cognitive neuroscience research has shown that these are not mutually exclusive alternatives. Instead, attention can operate at different stages depending on the nature of the stimuli and task. Moreover, recent research has moved beyond coarse distinctions between perceptual and postperceptual stages and is now examining how attention operates in different ways within the dozens of different areas of the visual processing pathway. This new research is reviewed by Kastner, McMains, and Beck in chapter 13, on neuroimaging research, and by Maunsell in chapter 19, on single-unit recordings.

Research on the implementation of attention has also become more precise in identifying the nature of attentional modulations of visual sensory responses. Two major principles have now been supported by many forms of converging evidence. First, attentional selection depends strongly on the degree of competition. As described in the chapters by Kastner, McMains, and Beck and by Maunsell, attention has its strongest effects when task-irrelevant information competes with the processing of task-relevant information. For example, attention effects are often observed to increase as information moves into higher stages of the visual processing pathway, where receptive fields are larger and are therefore likely to contain task-irrelevant objects as well as task-relevant objects. A second major principle is that attention often operates as a gain control, increasing the effective contrast for attended stimuli without changing sensory tuning curves. This point has been made most clearly in the psychophysical and single-unit studies reviewed by Maunsell and in the event-related potential (ERP) studies reviewed in chapter 15 by Hopf, Heinze, Schoenfeld, and Hillyard. We are also now beginning to understand how these effects may arise from the microcircuitry and temporal dynamics of visual cortex, as reviewed in chapter 20 by Womelsdorf and Fries on the role of neural synchronization in attention.

The control of attention is also described in considerable detail in this section. Much of this work takes place within the concept of a network of frontal, parietal, and superior temporal areas that control the operation of attention. The chapters in this section make the case for specializations in attentional control within these interconnected brain regions. In chapter 17, Karnath describes the detailed anatomy of a network of areas surrounding the Sylvian fissure involved in the control of attention, noting an interesting correspondence between right-hemisphere areas that are involved in attentional control and left-hemisphere areas that are involved in language and imitation. Chapter 14 by Corbetta, Sylvester, and Shulman complements this, describing how attentional control networks receive information from subcortical systems involved in affect and motivation and how different dorsal and ventral frontal-parietal networks are involved in the control of sensory and motor systems. Mangun, Saron, and Walsh describe how frontal and parietal attentional control networks interact with frontal cortical systems, such as the anterior cingulate cortex, that are involved in conflict detection, error monitoring, and online behavioral adjustments (chapter 16).

Research on the control of attention has also focused on the elementary "units" of selection. That is, does attention select spatial locations, nonspatial features, or whole objects? As reviewed in Treisman's chapter, this topic has been the source of much dispute in the cognitive literature for more than two decades, but cognitive neuroscience research has shown that these are not mutually exclusive alternatives. Attention can be allocated to the features of objects, to the locations occupied by objects, and directly to the objects themselves. This finding can be seen in neuroimaging studies (as described by Beck and Kastner and by Corbetta et al.), in single-unit studies (as described by Maunsell), and in ERP studies (as described by Hopf et al.). Perhaps the most impressive evidence, however, comes from the lesion patients described by Robertson in chapter 18. When spatial processing is severely disrupted in these patients, it is still possible to see clear evidence of intact feature-based and object-based attention.

Now that we have reviewed the major themes discussed by the chapters in this section, we would like to highlight three areas in which substantial progress has been made since the last edition of this book. First, although we have known for decades that a broad set of frontal, parietal, and subcortical areas play a role in the control of attention, the chapters in this section provide a much more detailed description of how these areas work both independently and in concert to control attention. New research is also looking at the microcircuitry of attention, revealing how different subclasses of neurons within an area are modulated by the operation of attention.

Second, the new research described in this section is beginning to reveal the details of how frontal and parietal areas can control the operation of attention within sensory areas. Top-down control signals have now been measured in a broad variety of attention tasks, and stimulation techniques have been used to show that activity within control areas can directly modulate sensory responses within visual cortex. A major unknown, however, is how these signals lead to the changes in neural synchrony that now appear to play an important role in the implementation of selection.

Finally, the chapters in this section describe new insights into the allocation of attention to nonspatial features, such as color and direction of motion. Recent studies have shown that attention can be directed to specific feature values across the visual field, and not just at attended locations. Indeed, the operation of feature-based attention sometimes precedes and guides the allocation of space-based attention. We expect that the next five years will lead to new insights into how these different varieties of attention work together in the service of perception and behavior.

12 Attention: Theoretical and Psychological Perspectives

ANNE TREISMAN

ABSTRACT This chapter reviews research on attention using behavioral and psychological methods. It attempts to illustrate what was learned through these tools alone and what is gained when tools from cognitive neuroscience are added. The psychological approaches defined many of the theoretical issues, such as the nature of the overloads that make attention necessary, the level of selection, the method of selection (enhancement of attended stimuli or suppression of unattended ones), the targets of selection (locations, objects, or attributes), the ways in which attention is controlled, and the role of attention in solving the feature-binding problem. Psychology also developed many of the paradigms used to probe the underlying mechanisms that are now being confirmed by converging evidence from brain imaging and from studies of brain-damaged patients. Theories of attention have evolved from the early sequential "pipeline" model of processing to a more flexible and interactive model with parallel streams specializing in different forms of perceptual analysis, iterative cycles of processing, and reentry to earlier levels. Attentional selection takes many forms and applies at many levels. We learn as much from exploring the constraints on flexibility—what cannot be done—as from discovering what can.

While watching a movie of a basketball game and counting the passes made by one of the teams, participants completely miss seeing a large black gorilla walk through the game, even though it is clearly visible if attended to (Simons & Chabris, 1999; see also Neisser & Becklen, 1975). Why should we still be interested in purely psychological studies like this one? If, as Minsky said, the mind is what the brain does, then that is what, as psychologists, we are interested in, and it would be foolish not to use the tools from neuroscience. However, brain-imaging data and findings with neurological patients depend critically for their interpretation on the designs of the behavioral tasks being performed. We can directly observe actions, or we can measure brain activation, but by putting them together we further constrain the possible theories. This chapter is intended to set the scene, both historically and conceptually, for the subsequent chapters exploring neuroscientific approaches to attention in more detail.

The traditional questions in attention research mostly started as di- or trichotomies: "Is it *x*, or *y*, or *z*?" For example, does attention select the relevant stimuli early or late in perceptual processing? The answer typically proves to be "both or all the above." Simplistic questions have evolved into attempts to specify *when* each answer applies and *why*. I select eight such issues to discuss here, using mostly psychological methods and bringing in neural evidence where it can decide questions that otherwise could not be answered. I also introduce many experimental paradigms that have been used to study attention. The goal has been to "bottle" the wide range of everyday phenomena encompassed by the label "attention" and bring them into the controlled conditions necessary for scientific study.

ANNE TREISMAN Psychology Department, Princeton University, Princeton, New Jersey

Why is attention limited?

As the gorilla example that opened this chapter suggests, attention seems to be severely limited. Other examples abound, as we show in later sections. We typically see only four items in a brief visual flash (Woodworth, 1938). We can follow the content of only one auditory message at a time (Broadbent, 1958). We can track only four moving circles among other identical circles moving in random directions (Pylyshyn & Storm, 1988). Why do these limits arise? There are three general ideas about their nature, and all could play a part.

STRUCTURAL INTERFERENCE One hypothesis is that limits arise only when two concurrent tasks use the same specialized subsystems. Proponents compared the interference between tasks that seemed likely to share common mechanisms and tasks that did not—for example, both were speech shadowing or one was piano playing (Allport, Antonis, & Reynolds, 1972), both were visual or one was auditory (Treisman & Davies, 1973), both used verbal rehearsal or one used imagery (Brooks, 1968). The results clearly showed more interference between tasks that were more similar. When they were sufficiently different, they were sometimes combined without impairment.

Different attributes like color, motion, and shape are processed by at least partially separate systems (e.g., Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991). Thus the structural interference view predicts little difficulty

in registering different properties of the same object, and this is what is found. Interference may still arise at the response level if the different attributes evoke conflicting responses, as in the task of naming the colored inks in which different color names are written (Stroop, 1935). The reason why attention is ineffective in tasks like the Stroop may be that the brain is forced to use whatever discriminative systems it has available *unless* these are fully occupied with other stimuli (Treisman, 1969).

Certain stimuli may have privileged access to attention, bypassing the structural limits: In a dichotic listening task, participants' own names sometimes broke through from the unattended message (Moray, 1959). Emotional stimuli, like a snake (Ohman, Flykt, & Esteves, 2001) or an angry face, are more likely to be seen than neutral stimuli (Eastwood, Smilek, & Merikle, 2001; for a review see Vuilleumier, 2005). The advantage extends also to guns and other nonevolutionary stimuli, suggesting that attention can be drawn to learned categories of fearful stimuli (Blanchette, 2006). Emotional stimuli may directly activate a separate pathway to the ventral prefrontal cortex and the amygdala (e.g., Yamasaki, LaBar, & McCarthy, 2002), although, contrary to some prior claims, some attentional resources are still needed for their detection (Pessoa, Kastner, & Ungerleider, 2002).

General Resources There are also more general limits to attention. Kahneman (1973) argued for a limited pool of resources or "effort." He showed that a secondary task (monitoring a stream of visual letters) was impaired when combined with a primary task of adding one to each of a string of auditory digits, although these two tasks are unlikely to share the same brain systems. Kahneman used the size of the pupil as an online index of effort, having previously shown that it correlates closely with difficulty across a wide range of tasks. Interference with visual letter detection was maximal when the memory load was highest and effort, as indexed by the pupil, was at a peak (see figure 12.1).

Purely psychological studies are handicapped in determining what information is extracted from unattended messages by the fact that observable responses are needed as evidence, so that limits could arise from our inability to carry out simultaneous actions or to remember stimuli that we did in fact observe. Brain imaging allows us to monitor the incidental processing of unattended stimuli as they are presented, and it may give us more sensitive indications of where the limits arise. Results have cast additional doubt on claims that only structural interference matters. Rees, Frith, and Lavie (1997) observed fMRI activity in area MT produced by irrelevant moving dots surrounding a central, task-relevant word. Two tasks differing in difficulty were used to assess the effects of load in processing the central word: case discrimination (easy) or detecting a bisyllabic word (more difficult). Discriminating phonology is unlikely to involve area MT, yet fMRI activation to the irrelevant dots was reduced during the more difficult word task. Thus attention limits appear between two very different forms of visual perception. However, when auditory words replaced the visual ones, task difficulty in the auditory word task had no impact on visual activation of area MT. Psychological tests provided converging evidence: With visual words, the difficult task reduced the motion aftereffect generated by the irrelevant dots, whereas with auditory words it was unaffected. Resources are at least partly shared across very different tasks within vision, suggesting resource limits rather than structural interference within the visual modality, but may be separate across different modalities. An interesting exception concerns spatial coding, where a shared representation of space may create some overlap in resources (e.g., Spence & Driver, 1996).

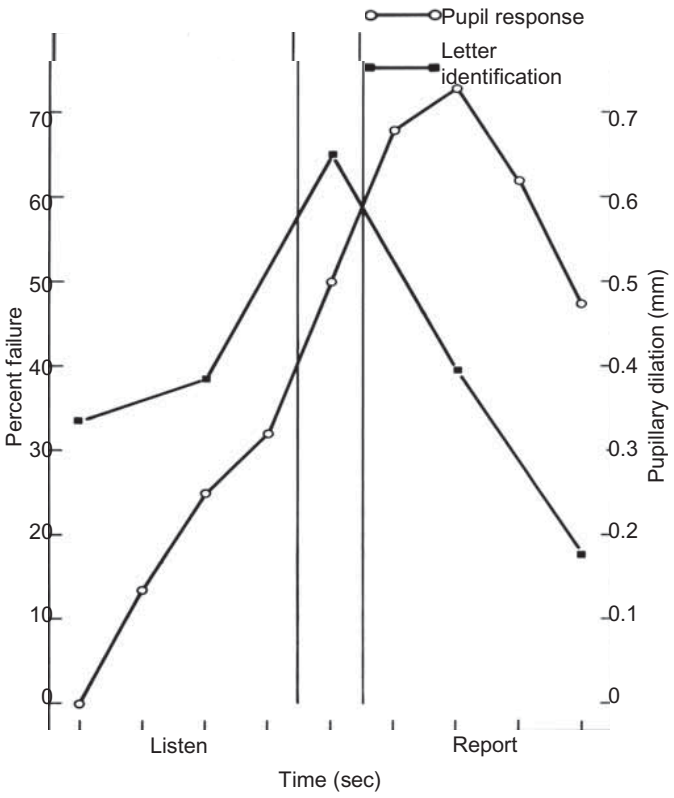


Figure 12.1 A measure of perceptual deficit and the pupillary response to a digit-transformation task. Black symbols show percent missed letters in a rapid visual sequence while participants listened to four digits, adding one to each and reporting the results (at rate of 1 per second). Errors increase with each extra digit at intake; they are highest at the time when participants are doing the mental addition; and they decrease as each transformed digit is reported. Open symbols show the size of the pupil, reflecting the amount of effort or resources being used. Note that the pupil index has about a 2-second lag behind the mental processing. (Modified with permission from Kahneman, 1973.)

Behavioral Coherence The premotor theory of attention suggests that attention is simply a preparation for response, selecting the goal of an intended action (Rizzolatti, Riggio, Dascola, & Umilta, 1987). Attention is facilitated when the actions afforded by the stimulus are compatible with the response required (e.g., Craighero, Fadiga, Rizzolatti, & Umilta, 1999; Tucker & Ellis, 1998). Hand location can facilitate detection of targets near the hand (Reed, Grubb, & Steele, 2006). Spatial attention normally follows a saccadic eye movement and can be triggered by subliminal stimulation of neurons that control that saccade (Moore & Armstrong, 2003). Attentional systems differ for space that is within reach and space that is beyond it (Ladavas, 2002). Actions may themselves affect the way attention is deployed. If a salient target like a unique color need only be detected, it "pops out" of the display; but if the goal is to touch the object, focused attention is required (Song & Nakayama, 2006). Although motor performance clearly interacts with attention, it seems unlikely that intended actions are the only limits to attention. Even when we passively watch stimuli go by (e.g., at the movies), we do select a subset of the information that reaches the senses.

When Limits Disappear Attention limits are typically found in unpracticed tasks, but most attention-demanding tasks can, with extensive practice, be automatized—made independent of attention. Search for target letters among other random letters initially demands attention and gives steep slopes of latencies against display size. But after weeks of practice with consistent target-distractor mapping, the slopes become flat (Schneider & Shiffrin, 1977). Two verbal tasks that are both similar and demanding (reading for comprehension and writing to dictation) can initially be done only in alternation, but after weeks of practice, they are efficiently combined (Spelke, Hirst, & Neisser, 1976). The only remaining constraint is between tasks that required combining words in sentences, for example, deciding whether the passage makes sense semantically. Theories of attention limits must also account for how these limits can, in many cases, disappear with practice.

What does attention select?

Traditionally the debate has centered on three possible goals of selection—locations, objects, and attributes. To these, we can add selection of scale and selection of moments of time. The answer again seems to be "all the above." Spatial selection is implied when facilitation spreads to separate, unrelated objects in the neighborhood of an attended target, shown by behavioral improvement (Hoffman & Nelson, 1981), ERPs (Mangun & Hillyard, 1988), and fMRI activation (Downing, Liu, & Kanwisher, 2001). For example, Downing and colleagues showed that attending to one of two separately located colored ovals increased the fMRI activation produced by either a house or a face stimulus that shared the same location.

A number of findings also favor object selection. When overlapping objects share the same general location, like the basketball game and the gorilla described earlier, attention can still be very efficient. Selection may be guided by properties of the attended object, perhaps a color or a range of spatial frequencies, or simply by the collinearity and spatial continuity of its contours. Attention spreads more easily within than between objects (Duncan, 1984; Egly, Driver, & Rafal, 1994; Tipper & Behrmann, 1996; see figure 12.2). Patients with neglect due to parietal damage who are oblivious to the left side of space often also neglect the left side of objects (e.g., Halligan & Marshall, 1994). A dramatic

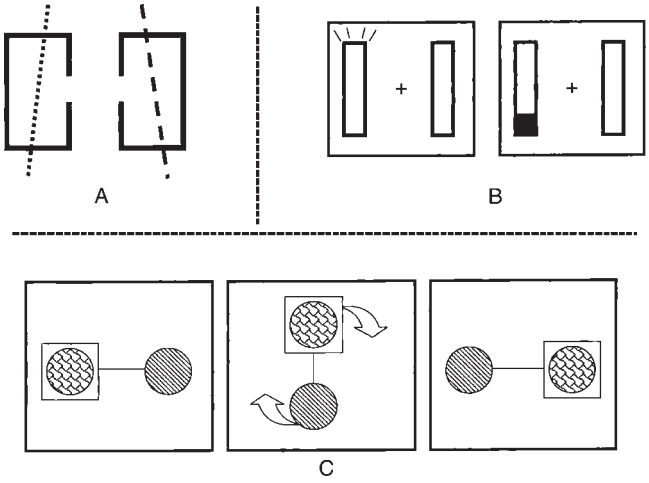


Figure 12.2 Three experiments suggesting that attention is allocated to perceptual objects. (*A*) Two examples of stimuli used in the experiment. Only one was presented on any trial. Participants were better at making two decisions on a single object (e.g., the orientation and texture of the line) than on two objects (e.g., the orientation of the line and the gap side of the rectangle), even though the two objects share the same location. (From Duncan, 1984, with permission.) (*B*) Two bars are presented, and one end of one bar brightens temporarily. Then a target (indicated by the black area) is presented either at the cued end of the cued object, or at the opposite end (as in the example shown), or in the other object at the same spatial distance from the cue. Responses are fastest at the cued end, but they are also faster at the other end of the cued object than in the other object, indicating that a within-object shift of attention is faster than a between-objects shift. (Modified with permission from Egly, Driver, & Rafal, 1994.) (*C*) Schematic illustration of the impact of left-sided neglect in the rotating objects experiment. When shown two objects, neglect patients with right parietal damage often see the one on the right and fail to see the one on the left. If they watch the objects rotate, however, they may continue to neglect the one that was originally on the left, even when it moves into the right field, suggesting that attention is allocated to the object rather than to its location. The shaded area reflects the part of the display neglected by the patients. (Modified with permission from Tipper & Behrmann, 1996.)

demonstration of object-based selection with *no* distinguishing properties involves tracking a subset of identical moving dots (Pylyshyn & Storm, 1988). Participants are shown, for example, eight identical dots, of which four briefly brighten, indicating that they are the targets. All eight dots then move on random, independent paths, and participants attempt to track the four targets. Selection here must be based exclusively on spatiotemporal continuity, since after the initial cues nothing else differentiates targets from distractors. Scholl, Pylyshyn, and Feldman (2001) further defined what counts as a selectable object, showing that the ends of lines whose other ends are to be ignored are impossible to track. Pylyshyn proposed that selection is maintained through preattentive indices (FINSTs) attached to the cued targets. These give direct access to the objects to which they are attached, and finding them should therefore not need attention. However, the task is clearly subject to attention limits, showing a decrement with dual tasks (Treisman, 1993). Wilson and I showed that tracking was impaired when participants were also asked to note changes in binding of color and texture in the borders around the display.

Attributes may also be units of attentional selection. For example, we may attend to motion and ignore color and shape (Corbetta et al., 1991). However, selection of one attribute within an object is often less efficient than other forms of selection. There is a strong bias to attend to objects as wholes. The Stroop task illustrates a failure to reject a word while attending to its color. Can we further narrow attention to select a particular feature within an attribute (e.g., red within the attribute of color)? Serences and Boynton (2007) showed that facilitation of a particular feature could spread across the visual field, even in the absence of a stimulus. Observers can track one feature of an object as it changes over time, so long as the changes are gradual (Blaser, Pylyshyn, & Holcombe, 2000). However, they could track two changing features at once only if these characterized the same object. Again feature selection was mediated by attention to an object as a whole.

When irrelevant attributes produce no behavioral interference with the relevant task, brain imaging is needed to test whether they are truly suppressed. O'Craven, Downing, and Kanwisher (1999) used overlapping face and place stimuli and asked participants to report either the direction of the moving picture or the location (slightly shifted from center) of the static one. The picture with the relevant attribute produced increased activation in the brain area specialized for its (irrelevant) category—PPA for houses and FFA for faces. Selection of one attribute here also resulted in enhanced attention to the object as a whole. The picture that emerges is that attention can be biased to select any of the three candidates, locations, objects, and attributes, but that object selection either takes precedence over feature selection or mediates it.

Selection of scale with hierarchically organized stimuli (e.g., global letters made of local letters) shows many similar attentional phenomena. For example, it can take time to reset attention from one level to another (Ward, 1982), and performance at one level deteriorates when attention is cued to the other (Bulakowski, Bressler, & Whitney, 2007). Navon (1977) found faster responses to the global level ("global precedence"), although this preference is modulated by the density and size of the local elements (Kimchi, 1992). Ivry and Robertson (1998) suggest that the attentional bias is relative rather than absolute, selecting the *higher* or the *lower* spatial frequencies in any given display. They link it to a cerebral asymmetry favoring higher frequencies in the left hemisphere and lower in the right. Data from neurological patients support this specialization (Robertson & Lamb, 1991).

Finally selection of moments in time is demonstrated by comparing performance with and without a temporal warning. In Posner's typology of attention, measured by his Attention Network Test (ANT; Fan, McCandliss, Sommer, Raz, & Posner, 2002), this is one of the three components of attention—alerting, orienting, and executive control. Alerting, cued by a temporal warning, considerably speeds reaction times.

How does attention select?

Does attention enhance relevant items, inhibit irrelevant items, or change the tuning of selected neurons? This question is difficult to answer because it is unclear what the baseline should be. We cannot be in a state of *no* attentional deployment. The alternatives to focused attention are inattention (i.e., focused attention to another object) or divided attention to two or more objects, which allows only a relative measure of facilitation. Probing unattended objects behaviorally is self-defeating, although an indirect measure of reduced interference can suggest inhibition. Another solution is to use neural measures of response to unattended stimuli.

Facilitation Focused attention is generally found to facilitate the processing of selected signals relative to the divided attention baseline, improving the accuracy or latency of response (Posner, 1980). There is some disagreement about the form the facilitation takes—an increase in signal-to-noise ratio, a narrower tuning to sharpen discrimination, or a shift in criterion (see chapter 19 by Maunsell, this volume, for evidence from neural recordings). Prinzmetal, Nwachuku, Bodanski, Blumenfeld, and Shimizu (1997) and Prinzmetal, Amiri, Allen, and Edwards (1998) investigated the phenomenology of attention, showing that it reduces the variance but leaves the perceived intensity unchanged. Signal detection theory separates effects on sensitivity (d')

and on decision criteria. Shaw (1984) found that in detection of luminance increments, attention load affected only the criterion, but in a letter localization task it also affected *d′*. However, Hawkins and colleagues (1990) found that spatial cuing affected both *d′* and criterion in a luminance detection task.

A related question, best answered with neural measures, is whether attention produces a multiplicative effect on the signal (gain control) or simply changes the baseline activity on which the signal is superimposed. Hillyard, Vogel, and Luck (1998) concluded that the gain control model fits best in early spatial selection. But there is also evidence in some conditions for changes in baseline activity (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Chawla, Rees, & Friston, 1999).

Inhibition Whether inhibition is invoked in an attention task may depend on how distracting the irrelevant stimuli would otherwise be. More active suppression is needed when the target and distractors are superimposed rather than spatially separated. Participants name one of two superimposed pictures more slowly when the currently relevant picture was the irrelevant one on the previous trial, suggesting that it was inhibited when it was irrelevant and the inhibition then had to be removed (Tipper, 1985). Thus this negative priming paradigm may show aftereffects of inhibition. A reduction in Stroop interference demonstrates the online inhibition of an irrelevant object, not just the aftereffects of inattention (Wühr & Frings, 2008; see figure 12.3).

Inhibition may also be used to prevent rechecking the same locations or stimuli that have already proved fruitless. Thus responses are slower at locations that have previously received attention, an effect known as inhibition of return (Posner & Snyder, 1975). Probe items in search get slightly slower responses when they appear in locations previously occupied by nontarget items (Cepeda, Cave, Bichot, & Kim, 1998; Klein & MacInnes, 1999). In the Marking Paradigm (Watson & Humphreys, 1997), a subset of distractors is shown in advance of the full search display. This procedure eliminates their contribution to search latencies, producing efficient feature search of just the items that appear in the final display instead of what would otherwise be a slower search for a conjunction target.

Functional MRI measures offer more direct evidence of inhibition, suggesting that attention to a stimulus at the fovea strongly suppresses baseline activity in brain areas responding to other spatial locations (Smith, Singh, & Greenlee, 2000). Event-related potential (ERP) differences between attended and unattended items show both inhibition of irrelevant items (shown in the P1 ERP component) and facilitation of relevant ones (shown in the N1 component) (Luck et al., 1994). When participants must bind features to identify the target, both a P1 and an N1 are shown, whereas when the presence of a color is sufficient, only the N1 (facilitation) effect remains (Luck & Hillyard, 1995). Again, inhibition is used only when distractors would otherwise cause interference.

Changes of Tuning or Selectivity Receptive field sizes can change, shrinking with attention to give finer selectivity (Moran & Desimone, 1985). Selectivity to particular features can also be sharpened. For example, when participants attended to the direction in which objects were rotated, the selectivity of fMRI in area LOC to orientation differences was increased relative to when they attended to the color of a central dot (Murray & Wojciulik, 2004). (Again, see chapter 19 by Maunsell for evidence from neural recordings.)

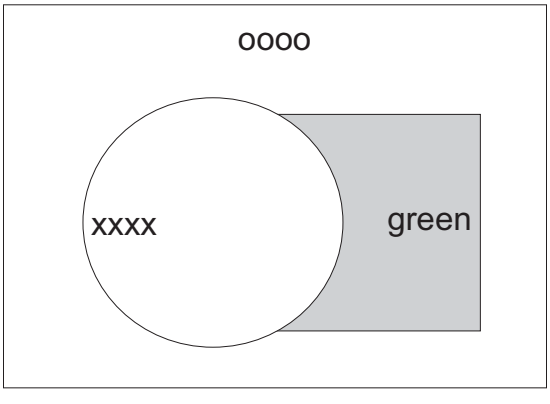


Figure 12.3 The task is to name the color of the square shape (which was yellow). Stroop interference is greater from the word "green" inside the relevant object than it is if the word is presented in the background where the *O*'s are shown, and it is reduced further when presented in the irrelevant red circle, where the *X*'s are shown in the figure. (Modified with permission from Wühr & Frings, 2008.)

Does attention act early or late?

In the early days of attention research, information processing was seen as a pipeline of successive stages, the output of each becoming the input of the next, with information of increasing complexity abstracted at each level. Attention could potentially select between outputs at any level to determine which should be passed on to the next. This model has been replaced by a more interactive system with reentry to early levels and extensive lateral communication between separate parallel streams of analysis, dealing with different types of information—"what?" in the ventral versus "where?" in the dorsal areas (Ungerleider & Mishkin, 1982), or objects and events for conscious representation in the ventral pathway and the online control of actions in the dorsal pathway (Milner & Goodale, 1995). Within each pathway, selection can occur at various levels, depending on the task,

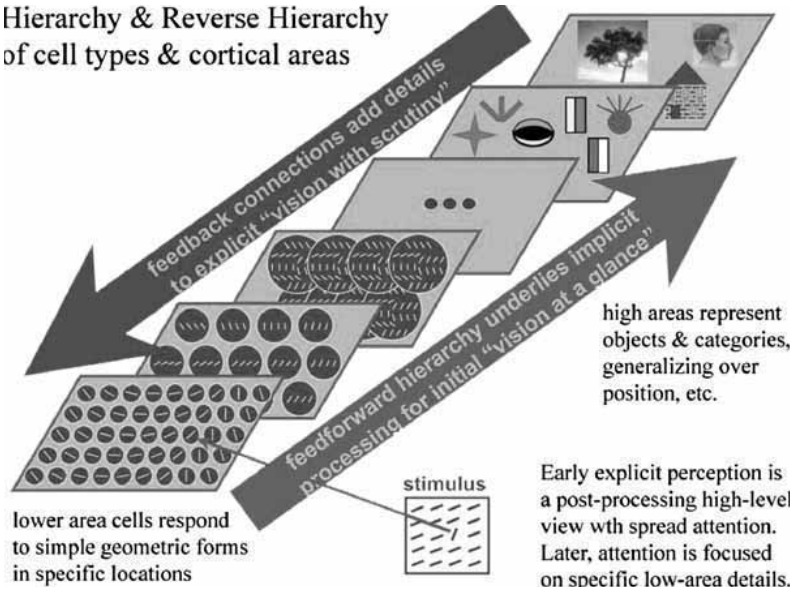


FIGURE 12.4 The reverse hierarchy theory. An initial fast-forward sweep through the visual hierarchy yields early access to high-level properties from large receptive fields. This can be followed by reentry to earlier areas, as required by the task. (Reprinted from Hochstein & Ahissar, 2002.)

the load, and the degree to which concurrent tasks engage the same subsystems. A recent framework that captures this flexibility is the reverse hierarchy theory of Hochstein and Ahissar (2002), in which an initial feedforward sweep through the sequence of visual areas takes place automatically, followed by optional controlled processing that may return to lower areas as required by the task (see figure 12.4). Access to awareness is initially at the highest levels of representation where receptive fields are large and discrimination is categorical.

Without assuming that the two must be correlated in a fixed order, we can still ask about either the level or the time at which selection is made. The early-late dichotomy was actually always a "straw" question. Proponents of early selection did not deny that selection could also occur late. The real question was "*Can* attention act early (Broadbent, 1958), or is all perceptual processing automatic, with attention selecting only at the level of memory and response?" (Deutsch & Deutsch, 1963). On one hand, proponents of late selection argued that attention limits are determined by decision effects alone: More stimuli lead to increased uncertainty and increased chances that noise will exceed a response criterion (e.g., Bundesen, 1990; Kinchla, 1974; Palmer, 1995). On the other hand, behavioral tests showed that selection based on properties that are processed early (simple physical characteristics like location, color, auditory pitch) was more efficient than selection based on properties presumably processed only later (semantic content, abstract categories). This conclusion was true even when the load on responses and memory was minimized (e.g., Treisman & Riley, 1969).

In the so-called psychological refractory period, attention limits seem to arise late. When separate speeded responses are required to two stimuli presented in close succession, the response to the second is typically delayed, reflecting an attentional bottleneck (Welford, 1952). Pashler (1993, 1994) used evidence of underadditivity of factors contributing to the two reaction times to locate the point at which overlap in processing becomes impossible. He found convincing evidence that the bottleneck arises not in perception but in central decision and response selection (see figure 12.5). The fact that attention limits *can* and sometimes do arise at late stages does not refute the claim that they can also act early.

A coherent account relates the level of selection to the level at which the potential overload occurs. If perception is demanding, selection needs to be early, whereas if the perceptual load is low, early selection may be not only unnecessary but actually impossible. Lavie and Tsal (1994) and Lavie (1995) showed that interference from a flanking distractor decreased, and, by inference, early selection efficiency increased, as the attended task became more difficult (see figure 12.6). But if the load arises in the control systems that direct attention, high load may *reduce* the efficiency of early selection and increase the effects of irrelevant stimuli. Using a dual task where participants were to remember the order of four digits, presented either in random orders (high working memory load) or in a fixed regular order (low load), while classifying famous names printed over irrelevant distractor faces, de Fockert, Rees, Frith, and Lavie (2001) found that incongruent faces produced more interference in the *high* load condition, presumably because working memory

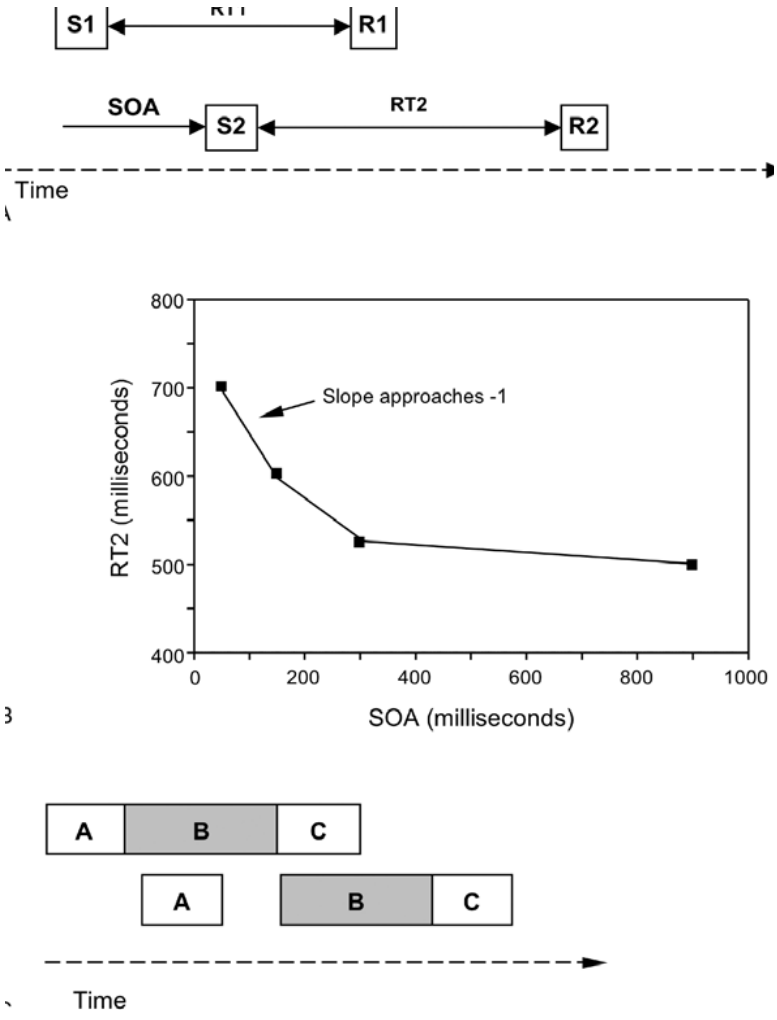


FIGURE 12.5 Psychological refractory period (*A*) Objective sequence of events (S1, stimulus 1; R1, response 1; SOA, stimulus onset asynchrony). (*B*) Observed reaction time to second stimulus is delayed as the interval between the tasks is reduced. The slope approaches −1, indicating that (on average) the second response cannot be produced until a certain time after S1. (*C*) Hypothesized stages of mental processing: A, perceptual processing; B, central decision time; C, response programming time. Stage B forms a bottleneck where two separate decisions cannot overlap with each other. Other stages can operate in parallel. (From Pashler, 1994, with permission.)

involves the same frontal lobe executive system as attentional selection. Brain imaging provided converging evidence: Activation in the fusiform face area was higher when working memory load was high, making selection inefficient.

IMPLICIT PROCESSING In distinguishing the level at which selection is made, we must also distinguish implicit processing from explicit accessibility. In the 1960s and 1970s, it was often assumed that perceptual processing was fully reflected in conscious experience and that behavioral responses were a reliable guide to the information available. This assumption was challenged by an early finding (Corteen & Dunn, 1974): Shock-associated words in an unattended message produced a galvanic skin response without also being consciously detected. Since then, other examples of implicit processing have been documented, using both indirect behavioral evidence, such as priming, interference, or emotional responses, and direct neural measures of responses to unattended stimuli.

Patients with unilateral neglect due to a right parietal lesion often show indirect evidence that they have identified stimuli that they are unable to report (e.g., McGlinchey-Berroth, 1997). There are also many instances of implicit perception in normal participants. When participants judged the relative length of the arms of a cross, an additional unexpected stimulus was often simply not seen (Mack & Rock, 1998). Yet a subsequent word-completion task showed priming from visual words to which the participant was "inattentionally blind" (see figure 12.7). A smiling face and the participant's own name were among other stimuli that were also detected, presumably because of their subjective importance. Implicit processing of surprising complexity and

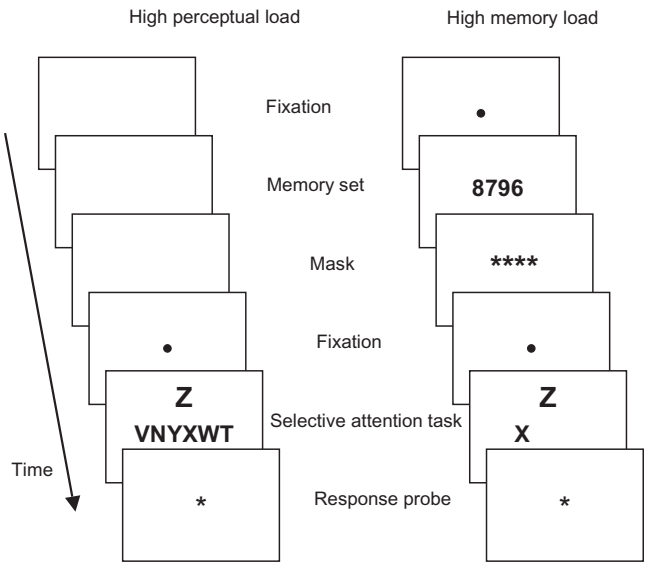


FIGURE 12.6 Two forms of attention load with different effects on performance. The task is to find the *X* or *Z* in the foveal letter string or single letter. On the left, the irrelevant flanking *Z* causes more interference when the relevant task involves a single letter (low load) relative to a string. On the right, there is always just one central letter, but there is an additional memory task with either a high load (4 random digits to remember) or a low load (4 digits in sequential order, e.g., 6789). Interference from the irrelevant flanker is higher in this case when the load on working memory is high. The reason suggested is that the control of attention depends on the same areas as working memory. (Modified with permission from Lavie & DeFockert, 2003.)

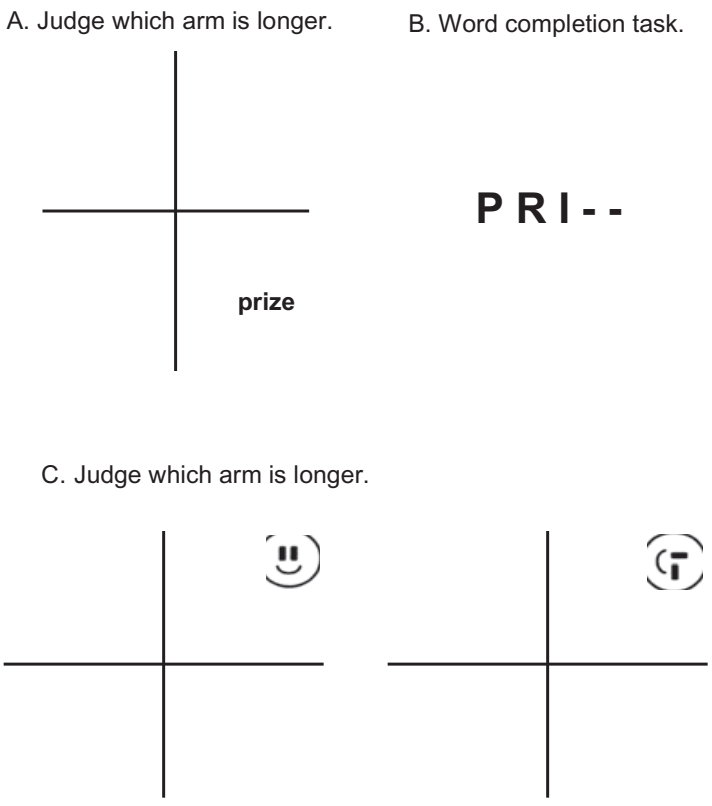


FIGURE 12.7 Inattentional blindness. (*A*, *B*) An experiment showing implicit processing of unattended words. On the first three trials participants are shown only a plus sign and they judge which arm is longer. On the fourth trial a word appears in one of the quadrants and is often not seen. However, in a subsequent word-completion task participants are more likely to respond with the unseen word, showing implicit priming from an unattended and undetected word. (*C*) A similar experiment in which the unexpected stimulus is a face. The smiling face is much more likely to be seen than the scrambled one or than a sad face, showing effects of meaning and emotion on implicit processing of unattended stimuli. (Modified from Mack & Rock, 1998.)

persistence was shown in a negative priming experiment (DeSchepper & Treisman, 1996). We found a slight delay in responding to novel, previously unattended nonsense shapes when they subsequently became relevant, even though explicit recognition memory was at chance (see figure 12.8). This negative priming, which sometimes lasted for days or weeks, was established in a single trial and was specific enough to discriminate between 260 different unfamiliar shapes.

Neural measures can show implicit processing even when behavioral responses do not. The attentional blink is shown when rapid serial visual presentation of two targets embedded in a string of distractors results in a reduced probability of detecting the second target within the following few items, as if the first occupies attention for several hundred milliseconds, blocking processing of the second (Shapiro & Raymond, 1994; see figure 12.9). Yet unexpected words presented during this "attentional blink" triggered an N400 ERP component, implying that they were identified at least to the level of semantic analysis (Luck, Vogel, & Shapiro, 1996). Yi, Woodman, Widders, Marois, and Chun (2004) presented a sequence of faces in the center and places in the surround. While attempting to detect a one-back repetition of a face, participants failed to notice repetitions of the unattended places, suggesting early selection. However, fMRI showed decreased activation to the repeated scenes in the parahippocampal place area. Adaptation to specific scenes implies that representations were formed and matched across trials despite never reaching conscious awareness. Again the fMRI effect was weaker than when the scenes received full attention, suggesting some reduction in early processing when attention is focused elsewhere.

Neural changes can specify the timing of attention effects. Functional MRI activation and single-unit changes occurring in anticipation of the stimulus have proved that attention can affect the baseline activity in specialized extrastriate areas even before the stimulus is presented—"early" indeed! (Chawla et al., 1999; Hopfinger, Buonocore, & Mangun, 2000; Kastner & Ungerleider, 2000). Early attenuation of unattended stimuli also appeared in ERP responses as early as 70–100 ms after stimulus onset (Van Voorhis & Hillyard, 1977). Anatomically low-level effects in an fMRI study

showed attention effects as early in the ascending visual pathways as the lateral geniculate (O'Connor, Fukui, Pinsk, & Kastner, 2002). Because of the multiple connections back from higher areas and the low temporal resolution of fMRI, it is ambiguous here whether attention affects a first pass through the visual hierarchy or acts only through reentrant connections. However, combined with temporal evidence from ERPs and single-neuron recordings, both reflecting early cortical activity, the occurrence of early, short-latency attentional modulation is now clearly established.

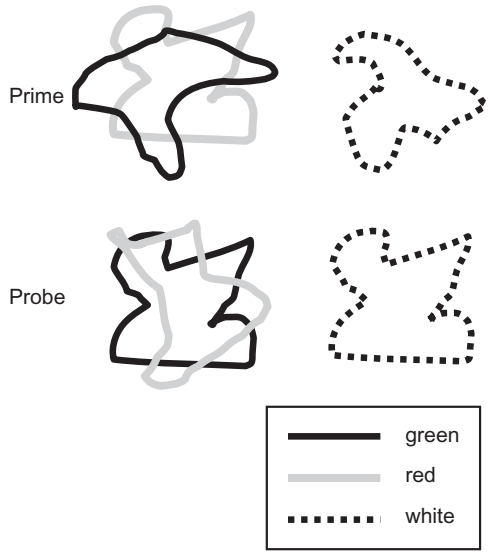
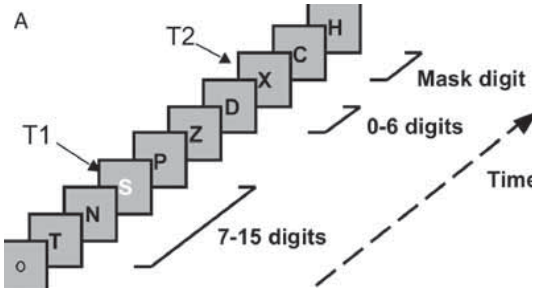


FIGURE 12.8 Negative priming. Participants judge whether the green shape on the left matches the white shape on the right while ignoring the red shape on the left. They have no explicit memory for the unattended shapes, yet when one reappears as the shape to be attended, responses are slightly slowed, as though it had previously been inhibited or labeled "irrelevant" and the label had to be cleared when the shape became relevant. This negative priming effect can last across hundreds of intervening trials and days or weeks of delay. (After DeSchepper & Treisman, 1996.)

Implicit processing of unattended items complicates the attentional story, suggesting that attention can block access to consciousness without blocking all forms of perceptual processing. The fact that this result can occur does not imply that it always does. The perceptual load was low in most experiments that showed implicit effects, and when it was raised, the implicit effects disappeared (Neumann & Deschepper, 1992; Rees et al., 1997). But we still need to explain why attention should limit conscious access when the perceptual load is low. Interference may be greater from consciously perceived objects than from implicitly distinguished stimuli.

The role of attention in feature binding

The visual system comprises many specialized areas coding different aspects of the scene. This modularity poses the binding problem—to specify how the information is recombined in the correct conjunctions—red shirt and blue pants rather than an illusory blue shirt. Behavioral results (Treisman & Gelade, 1980) suggest that we bind features by focusing attention on each object in turn. Evidence includes findings that when attention is prevented, binding errors or illusory conjunctions are frequently perceived (Treisman & Schmidt, 1982); a spatial precue helps detection of a conjunction much more than of a feature target (Treisman, 1988); boundaries between groups defined only by conjunctions are hard to detect, whereas those between features are easy (Treisman & Gelade, 1980); visual search depends on focused



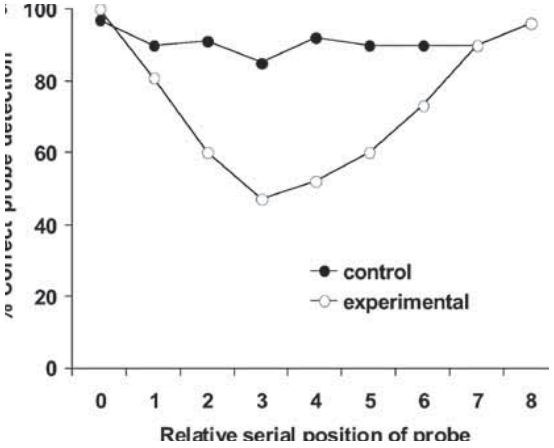


FIGURE 12.9 The attentional blink. (*A*) Participants monitor a rapid visual sequence for two different targets. For example, T1 might be a white letter in a black string, and T2 might be a letter X. (*B*) The open circles give the detection rates when participants do both tasks, and the black circles show performance on T2 in the control condition in which they ignore the first target. In the combined tasks, T2 is very likely to be missed if it occurs within a few hundred milliseconds of T1, suggesting that detecting a first target makes the participant refractory to detecting a second for the next few hundred milliseconds. (From Shapiro & Raymond, 1994, with permission.)

attention when the target is defined only by a conjunction of features (e.g., a green *T* among green *X*'s and brown *T*'s), whereas search for either of two disjunctive features (e.g., blue *or* curved) among the same distractors allows parallel search. In some cases, conjunction search for a known target with salient features can bypass the serial check, suggesting an additional mode of selection through feature grouping and guidance (Nakayama & Silverman, 1986; Treisman & Sato, 1990; Wolfe, Cave, & Franzel, 1989).

The theory was generally interpreted as predicting parallel search whenever a unique feature is available to mediate performance. This conclusion was correct, but the interpretation of "feature" as "any unidimensional difference" was not. When target and distractors differ on a single dimension, so that both activate the same detectors to different degrees, they can produce a wide range of search rates (Treisman & Gormican, 1988). Features are defined relative to a background: they are properties that activate neural detectors not also activated by surrounding elements. Converging tests include the following: Features are elements that can migrate to form illusory conjunctions, that mediate effective grouping and boundary detection, and that can be separately attended. Neural evidence from single units may also confirm candidates derived from behavioral criteria. There is evidence that new feature detectors can be established through learning (e.g., Freedman, Riesenhuber, Poggio, & Miller, 2002).

Support for feature integration theory came from a patient studied by Robertson, Treisman, Friedman-Hill, and Grabowecky (1997). RM had bilateral parietal damage from two successive strokes, showing the classic symptoms of Balint's syndrome: a failure to localize visual stimuli and an inability to see more than one object at a time—simultanagnosia. The theory predicted that he should also have difficulty binding features, since doing so requires focused attention to their shared location. Even with long exposures, when two letters were present, RM frequently saw the shape of one letter in the color of another. He was unable to search for conjunction targets, although, despite his severe simultanagnosia, he had no difficulty with feature search, detecting the presence of red in a display of blue letters, or a *Q* among *O*'s, with very few errors.

Further evidence from neuroimaging (Corbetta, Shulman, Miezin, & Petersen, 1995) showed that the same parietal areas were active in spatial attention switching and in binding, but not in feature search. Transcranial magnetic stimulation (TMS) to the parietal lobes selectively disrupts conjunction but not feature search (Ashbridge, Walsh, & Cowey, 1997). Luck and Hillyard (1995) showed ERP suppression specific to conjunction search. Recent studies comparing conjunction search with difficult feature search gave mixed results (Donner et al., 2002; Leonards, Sunaert, Van Hecke, & Orban, 2000; Wojciulik & Kanwisher, 1999). However, these are of questionable relevance, since the theory suggests that feature discrimination can also require focused attention when discriminability is low.

How is attention controlled?

Attention can be "captured" "bottom-up" by salient sensory events, particularly when they signal the onset of a new object (Yantis & Jonides, 1996). It may be attracted by events with emotional significance. But attention can also be endogenously controlled. Two general accounts have been proposed. One suggests a specialized attention network, controlling perceptual processing from the outside by amplifying relevant signals and/or attenuating irrelevant ones (Mesulam, 1981; Posner & Dehaene, 1994). The existence of such a control system is supported by several fMRI studies. D'Esposito and colleagues (1995) found that the dorsolateral prefrontal cortex and the anterior cingulate were activated only when two tasks were performed concurrently and not for either alone. Hopfinger and colleagues (2000), Corbetta, Kincade, Ollinger, McAvoy, and Shulman (2000), and Kastner and Ungerleider (2000) identify a network of areas, including FEF, SPL, SEF, IPL and MFG, involved in the external control of attention. Corbetta and colleagues additionally distinguish between voluntary control of attention in advance of the stimulus and the reorienting of attention when a target has been detected. The former seems to be controlled by the intraparietal sulcus, and the latter by the right temporoparietal junction (Corbetta et al.).

An alternative account sees selection as the emergent outcome of competition between neighboring objects within receptive fields. Reynolds, Chelazzi, and Desimone (1999) showed competitive interactions between single units in monkey ventral areas. The biased competition theory (Desimone & Duncan, 1995) combines the two accounts: top-down inputs from the control network bias the competition by adding to local activation of one of the competing objects. Different brain systems responding to the same object cooperate, ensuring that one object becomes dominant across multiple areas. It is not clear that purely behavioral evidence could distinguish biased competition from structural interference, which makes similar predictions about the specificity of interference. Thus the supporting evidence is mainly neural. Functional MRI studies showed weaker activation when four pictures were shown together for 250 ms than when they were shown successively for 250 ms each (Kastner & Ungerleider, 2000). Differences in stimulus duration were subsequently eliminated by presenting one picture in the upper visual field, either alone or together with three others in the lower field (Kastner et al., 2001). The added pictures reduced the activation produced by the target, but when attention was directed to the upper picture, full activation was restored. Competition arises

mainly within receptive fields. Since these increase in size with the level in the hierarchy, the further apart the stimuli, the higher the level of processing at which they compete (Kastner et al.). However, there may also be attention limits outside the classical receptive field. The early ERP effects of attention reflect selection between stimuli in different visual hemifields and therefore hemispheres of the brain (Van Voorhis & Hillyard, 1977). Another issue for the biased competition theory is to specify how local neurons "know" whether their activity is produced by parts of the same object, which should cooperate, or by different objects that should compete. This "knowledge" may require substantial top-down control of local competition.

Focused versus distributed attention

There is a paradox in attention research: Many studies show sharp limits such that only three or four objects can be tracked through space (Pylyshyn & Storm, 1988) or identified at a glance (Woodworth, 1938); change detection in an alternating pair of otherwise identical scenes is surprisingly difficult (Rensink, O'Regan, & Clark, 1997); and accurate binding depends on focused attention. Yet natural scenes can be rapidly and effortlessly monitored for semantic targets (Potter, 1975). While doing an attention-demanding task at the fovea, participants failed to discriminate which side of a peripheral circle was red, yet they easily detected an unknown animal target in a peripheral natural scene (Li, VanRullen, Koch, & Perona, 2002). This finding raises the question whether attention limits apply primarily to simplified laboratory stimuli, while in the natural world information is easily absorbed and understood. Can these contradictions be resolved?

Using natural scenes, Oliva and Torralba (2006) showed that the gist can be inferred from a combination of statistical properties. Chong and Treisman (2003) suggested that the global deployment of attention generates a statistical mode of processing. We confirmed a finding by Ariely (2001) that observers can accurately estimate the mean size of elements and showed that such estimating happens automatically when attention is distributed over the display (see figure 12.10). Combining these findings with the idea that the parallel intake of sets of diagnostic features could mediate detection of familiar objects even before those features are bound may resolve the paradox (Treisman, 2006). Participants, shown rapid sequences of natural scenes, were able to detect target animals quite well, while often being unable to specify which animal or where in the picture it appeared (Evans & Treisman, 2005). If two successive targets had to be identified, a severe attentional blink was incurred, but when the targets could simply be detected, the blink disappeared. If features must be bound to identify or locate a target, focused attention is required, but when only the gist is needed, statistical processing and parallel feature detection may provide sufficient information about most redundant natural scenes.

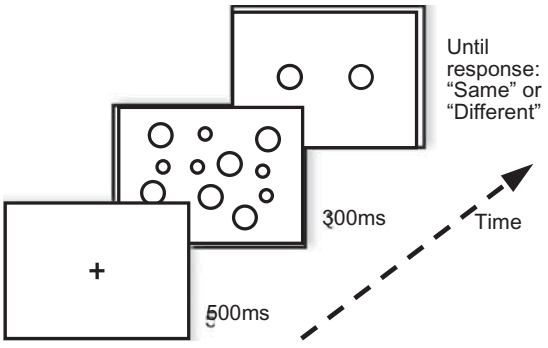


FIGURE 12.10 Implicit priming from the mean size of the previewed display. The task is to judge whether the two circles in the final display are the same size or different. Participants respond a little faster if one or both matches the mean size of the preceding prime display, and faster even than when they match one of the presented sizes. It seems that participants automatically compute the mean size of an array of circles. (Experiment described in Chong & Treisman, 2001.)

How does attention relate to consciousness?

Some theories equate attention and consciousness. This approach is probably misleading. Not everything that receives attention reaches awareness. We look at an ambiguous figure and experience only one interpretation. We may attend to a spatial location and show implicit priming without becoming aware that anything was there (Marcel, 1983a). Attention can facilitate unconscious perception in blindsight patients (Kentridge, Heywood, & Weiskrantz, 2004) and in normal participants (Kentridge, Nijboer, & Heywood, 2008). Attention limits appear even in unconscious perception: Kahneman and Chajczyk (1983) found "dilution" of interference when a neutral word was presented together with the irrelevant color name in a Stroop color-naming task. Bahrami, Lavie, and Rees (2007) found that load effects in a foveal task modulated V1 activation produced by unseen pictures of tools in the periphery.

Thus attention is not *sufficient* to ensure consciousness. Is it *necessary*? Can we be conscious of something that was not attended? There is an ambiguity here. I can be conscious of an unattended voice, without identifying the words that are spoken. Am I conscious of the stimulus? We are probably never conscious of every property, even of fully attended stimuli—for example, that this cat is smaller than the Eiffel tower. We become explicitly aware of just a small fraction of the possible propositions that could be formulated about an object or event that we are observing. With unattended objects, we lose those aspects for which capacity was overloaded, but we may retain some information.

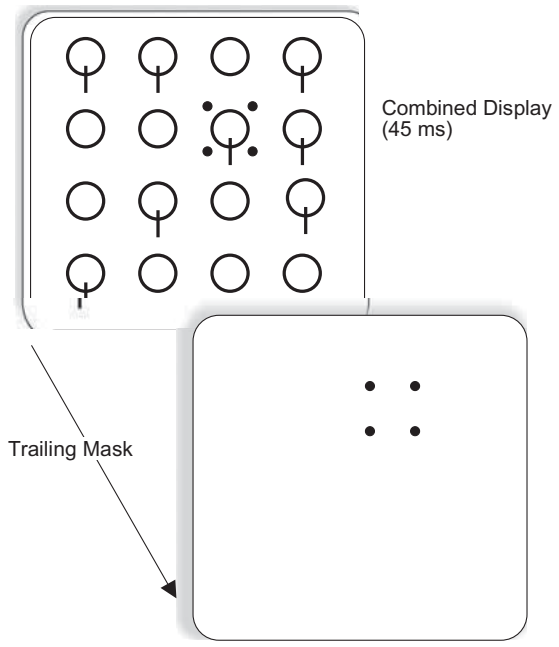


FIGURE 12.11 Object substitution masking. The display contains up to 16 rings, half of which have a vertical bar across the bottom. The target is singled out by four dots, as shown, which also serve as the mask. Observers indicate whether the target contains the vertical bar. The sequence begins with a combined display of the target, mask, and distractors for 45 ms and continues with a display of the mask alone for durations of 0, 45, 90, 135, or 180 ms. The four dots produce no masking if they end with the display, but if they continue after it disappears, they render the stimulus that they surround invisible. The suggestion is that visual processing will not reach conscious awareness unless a reentry check confirms the information extracted on a first pass through the visual system. In the case illustrated, the dots remain alone in the location of one of the *Q*'s and are substituted for it in conscious perception, whereas in the other locations there are no alternative stimuli to compete. (From Di Lollo, Enns, & Rensink, 2000.)

Attention, then, seems to be neither necessary nor sufficient for conscious awareness, although the two are normally highly correlated. What *is* necessary for conscious experience? The idea of reentry is much in the air these days. Several authors propose that the initial registration of stimuli consists of a rapid feedforward sweep through the visual areas without conscious awareness, then a possible return back to the early levels to check a tentative identification for selected elements against the sensory data (Damasio, 1989; Hochstein & Ahissar, 2002; Lamme & Roelfsema, 2000; Marcel, 1983b). Binding may also depend on reentry to early visual areas to ensure fine spatial resolution (Treisman, 1996). We become conscious of objects or events only if the initial match is confirmed. Supporting evidence comes from "object substitution" masking, in which a mask that begins at the same time as the target but outlasts it can render a stimulus invisible in a search array, even with no overlap of contours (Di Lollo, Enns, & Rensink, 2000; see figure 12.11). When the reentry check is made, only the mask remains, and it replaces the target in conscious experience. Related evidence comes by using TMS to V1 to erase the stimulus before the reentry check can be made, around 80 ms after onset (Walsh & Cowey, 1998; Lamme & Roelfsema, 2000).

Conclusions

Psychological studies posed many of the relevant questions, outlined possible mechanisms, and developed experimental paradigms to capture different aspects of what is meant by attention. The data provide constraints, ruling out many possible accounts. Neuroscience has added powerful tools to cast votes on issues that remained controversial, or sometimes to reframe the questions in ways that more closely match the way the brain functions.

Some suggest that attention acts primarily through biases on intrinsic competitive local interactions. Others see it arising primarily or only at the decision level, following parallel perceptual processing. Still others (myself included) suggest that conscious perception, detailed localization, and binding of features may depend on focused attention through reentrant pathways. An initial rapid pass through the visual hierarchy provides the global framework and gist of the scene and may prime target objects through the features that are detected. Attention is then focused back to early areas to allow a serial check of the initial rough bindings and to form the representations that are consciously experienced. The impact of neuroscience is obvious in these developments, but so is the ingenious and careful use of psychological paradigms to tease apart the mechanisms controlling our perception and action.

ACKNOWLEDGMENTS The work was supported by NIH grant number 2RO1 MH 058383-04A1; by the Israeli Binational Science Foundation, grant number 1000274; and by NIH grant R01 MH62331.

REFERENCES

ALLPORT, D. A., ANTONIS, B., & REYNOLDS, P. (1972). Division of attention—Disproof of single channel hypothesis. *Q. J. Exp. Psychol.*, *24*(May), 225–235.

ARIELY, D. (2001). Seeing sets: Representation by statistical properties. *Psychol. Sci.*, *12*(2), 157–162.

ASHBRIDGE, E., WALSH, V., & COWEY, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia*, *35*(8), 1121–1131.

BAHRAMI, B., LAVIE, N., & REES, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Curr. Biol.*, *17*(6), 509–513.

Blanchette, I. (2006). Snakes, spiders, guns, and syringes: How specific are evolutionary constraints on the detection of threatening stimuli? *Q. J. Exp. Psychol.*, *59*(8), 1484–1504.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, *408*(6809), 196–199.

Broadbent, D. E. (1958). *Perception and communication.* New York: Pergamon Press.

Brooks, L. R. (1968). Spatial and verbal components of act of recall. *Can. J. Psychol.*, *22*(5), 349–368.

Bulakowski, P. F., Bressler, D. W., & Whitney, D. (2007). Shared attentional resources for global and local motion processing. *J. Vis.*, *7*(10), 1–10.

Bundesen, C. (1990). A theory of visual-attention. *Psychol. Rev.*, *97*(4), 523–547.

Cepeda, N. J., Cave, K. R., Bichot, N. P., & Kim, M. S. (1998). Spatial selection via feature-driven inhibition of distractor locations. *Percept. Psychophys.*, *60*(5), 727–746.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nat. Neurosci.*, *2*(7), 671–676.

Chong, S. C., & Treisman, A. (2001). Representation of statistical properties. *J. Vis.*, *1*(3), 54.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vis. Res.*, *43*(4), 393–404.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.*, *3*(3), 292–297.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed—Functional-anatomy by positron emission tomography. *J. Neurosci.*, *11*(8), 2383–2402.

Corbetta, M., Shulman, G. L., Miezin, F. M., & Petersen, S. E. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science*, *270*(5237), 802–805.

Corteen, R. S., & Dunn, D. (1974). Shock-associated words in a nonattended message—Test for momentary awareness. *J. Exp. Psychol.*, *102*(6), 1143–1144.

Craighero, L., Fadiga, L., Rizzolatti, G., & Umilta, C. (1999). Action for perception: A motor-visual attentional effect. *J. Exp. Psychol. Hum. Percept. Perform.*, *25*(6), 1673–1692.

Damasio, A. R. (1989). Time-locked multiregional retroactivation—A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1–2), 25–62.

de Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, *291*(5509), 1803–1806.

DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: Implicit coding without attention. *J. Exp. Psychol. Learn. Mem. Cogn.*, *22*(1), 27–47.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual-attention. *Annu. Rev. Neurosci.*, *18*, 193–222.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working-memory. *Nature*, *378*(6554), 279–281.

Deutsch, J. A., & Deutsch, D. (1963). Attention—Some theoretical considerations. *Psychol. Rev.*, *70*(1), 80–90.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *J. Exp. Psychol. Gen.*, *129*(4), 481–507.

Donner, T. H., Kettermann, A., Diesch, E., Ostendorf, F., Villringer, A., & Brandt, S. A. (2002). Visual feature and conjunction searches of equal difficulty engage only partially overlapping frontoparietal networks. *Neuroimage*, *15*(1), 16–25.

Downing, P., Liu, J., & Kanwisher, N. (2001). Testing cognitive models of visual attention with fMRI and MEG. *Neuropsychologia*, *39*(12), 1329–1342.

Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.*, *113*(4), 501–517.

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Percept. Psychophys.*, *63*(6), 1004–1013.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual-attention between objects and locations—Evidence from normal and parietal lesion subjects. *J. Exp. Psychol. Gen.*, *123*(2), 161–177.

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *J. Exp. Psychol. Hum. Percept. Perform.*, *31*(6), 1476–1492.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *J. Cogn. Neurosci.*, *14*(3), 340–347.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2002). Visual categorization and the primate prefrontal cortex: Neurophysiology and behavior. *J. Neurophysiol.*, *88*(2), 929–941.

Halligan, P. W., & Marshall, J. C. (1994). Toward a principled explanation of unilateral neglect. *Cogn. Neuropsychol.*, *11*(2), 167–206.

Hawkins, H. L., Hillyard, S. A., Luck, S. J., Downing, C. J., Mouloua, M., & Woodward, D. P. (1990). Visual-attention modulates signal detectability. *J. Exp. Psychol. Hum. Percept. Perform.*, *16*(4), 802–811.

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. London B Biol. Sci.*, *353*(1373), 1257–1270.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.

Hoffman, J. E., & Nelson, B. (1981). Spatial selectivity in visual-search. *Percept. Psychophys.*, *30*(3), 283–290.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat. Neurosci.*, *3*(3), 284–291.

Ivry, R. B., & Robertson, L. C. (1998). *The two sides of perception.* Cambridge, MA: MIT Press.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading—Dilution of Stroop effects by color-irrelevant stimuli. *J. Exp. Psychol. Hum. Percept. Perform.*, *9*(4), 497–509.

Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *J. Neurophysiol.*, *86*(3), 1398–1411.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human

visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.*, *23*, 315–341.

Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (2004). Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia*, *42*(6), 831–835.

Kentridge, R. W., Nijboer, T. C. W., & Heywood, C. A. (2008). Attended but unseen: Visual attention is not sufficient for visual awareness. *Neuropsychologia*, *46*(3), 864–869.

Kimchi, R. (1992). Primacy of wholistic processing and global local paradigm—A critical-review. *Psychol. Bull.*, *112*(1), 24–38.

Kinchla, R. A. (1974). Detecting target elements in multi-element arrays—Confusability model. *Percept. Psychophys.*, *15*(1), 149–158.

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychol. Sci.*, *10*(4), 346–352.

Ladavas, E. (2002). Functional and dynamic properties of visual peripersonal space. *Trends Cogn. Sci.*, *6*(1), 17–22.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.*, *23*(11), 571–579.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *J. Exp. Psychol. Hum. Percept. Perform.*, *21*(3), 451–468.

Lavie, N., & DeFockert, J. W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Percept. Psychophys.*, *65*, 202–212.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual-attention. *Percept. Psychophys.*, *56*(2), 183–197.

Leonards, U., Sunaert, S., Van Hecke, P., & Orban, G. A. (2000). Attention mechanisms in visual search—An fMRI study. *J. Cogn. Neurosci.*, *12*, 61–75.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. USA*, *99*(14), 9596–9601.

Luck, S. J., & Hillyard, S. A. (1995). The role of attention in feature detection and conjunction discrimination—An electrophysiological analysis. *Int. J. Neurosci.*, *80*(1–4), 281–297.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cueing on luminance detectability—Psychophysical and electrophysiological evidence for early selection. *J. Exp. Psychol. Hum. Percept. Perform.*, *20*(4), 887–904.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616–618.

Mack, A., & Rock, I. (1998). *Inattentional blindness.* Cambridge, MA: MIT Press.

Mangun, G. R., & Hillyard, S. A. (1988). Spatial gradients of visual-attention—Behavioral and electrophysiological evidence. *Electroencephalogr. Clin. Neurophysiol.*, *70*(5), 417–428.

Marcel, A. J. (1983a). Conscious and unconscious perception—Experiments on visual masking and word recognition. *Cogn. Psych.*, *15*(2), 197–237.

Marcel, A. J. (1983b). Conscious and unconscious perception—An approach to the relations between phenomenal experience and perceptual processes. *Cogn. Psych.*, *15*(2), 238–300.

McGlinchey-Berroth, R. (1997). Visual information processing in hemispatial neglect. *Trends Cogn. Sci.*, *1*, 91–97.

Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Ann. Neurol.*, *10*(4), 309–325.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action.* Oxford, UK: Oxford University Press.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, *421*(6921), 370–373.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784.

Moray, N. (1959). Attention in dichotic-listening—Affective cues and the influence of instructions. *Q. J. Exp. Psychol.*, *11*(1), 56–60.

Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nat. Neurosci.*, *7*(1), 70–74.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*(6059), 264–265.

Navon, D. (1977). Forest before trees—Precedence of global features in visual-perception. *Cogn. Psych.*, *9*(3), 353–383.

Neisser, U., & Becklen, R. (1975). Selective looking—Attending to visually specified events. *Cogn. Psych.*, *7*(4), 480–494.

Neumann, E., & Deschepper, B. G. (1992). An inhibition-based fan effect—Evidence for an active suppression mechanism in selective attention. *Can. J. Psychol.*, *46*(1), 1–40.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nat. Neurosci.*, *5*(11), 1203–1209.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*(6753), 584–587.

Ohman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *J. Exp. Psychol. Gen.*, *130*(3), 466–478.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Prog. Brain Res.*, *155*, 23–36.

Palmer, J. (1995). Attention in visual-search—Distinguishing 4 causes of a set-size effect. *Curr. Dir. Psychol. Sci.*, *4*(4), 118–123.

Pashler, H. (1993). Dual-task interference and elementary mental mechanisms. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV* (pp. 245–264). Cambridge, MA: MIT Press.

Pashler, H. (1994). Dual-task interference in simple tasks—Data and theory. *Psychol. Bull.*, *116*(2), 220–244.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neutral and emotional stimuli. *Cogn. Brain Res.*, *15*(1), 31–45.

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.*, *32*(Feb.), 3–25.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends Neurosci.*, *17*(2), 75–79.

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabitt & S. Dornic (Eds.), *Attention and performance V* (pp. 669–682). New York: Academic Press.

Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965–966.

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention. 1. Color, location, orientation, and spatial frequency. *J. Exp. Psychol. Hum. Percept. Perform.*, *24*(1), 261–282.

Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., & Shimizu, N. (1997). The phenomenology of attention. 2. Brightness and contrast. *Consciousness Cogn.*, *6*(2–3), 372–412.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vis.*, *3*(3), 179–197.

Reed, C. L., Grubb, J. D., & Steele, C. (2006). Hands up: Attentional prioritization of space near the hand. *J. Exp. Psychol. Hum. Percept. Perform.*, *32*(1), 166–177.

Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science*, *278*(5343), 1616–1619.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychol. Sci.*, *8*(5), 368–373.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.*, *19*(5), 1736–1753.

Rizzolatti, G., Riggio, L., Dascola, I., & Umilta, C. (1987). Reorienting attention across the horizontal and vertical meridians—Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1A), 31–40.

Robertson, L., Treisman, A., Friedman-Hill, S., & Grabowecky, M. (1997). The interaction of spatial and object pathways: Evidence from Balint's syndrome. *J. Cogn. Neurosci.*, *9*(3), 295–317.

Robertson, L. C., & Lamb, M. R. (1991). Neuropsychological contributions to theories of part whole organization. *Cogn. Psych.*, *23*(2), 299–330.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information-processing. 1. Detection, search, and attention. *Psychol. Rev.*, *84*(1), 1–66.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*(1–2), 159–177.

Serences, J. T., & Boynton, G. M. (2007). The representation of behavioral choice for motion in human visual cortex. *J. Neurosci.*, *27*(47), 12893–12899.

Shapiro, K. L., & Raymond, J. E. (1994). Temporal allocation of visual attention: Inhibition or interference? In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory and language* (pp. 151–188). New York: Academic Press.

Shaw, M. L. (1984). Division of attention among spatial locations—A fundamental difference between detection of letters and detection of luminance increments. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* (pp. 109–121). Hillsdale, NJ: Erlbaum.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074.

Smith, A. T., Singh, K. D., & Greenlee, M. W. (2000). Attentional suppression of activity in the human visual cortex. *NeuroReport*, *11*(2), 271–277.

Song, J. H., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *J. Vis.*, *6*(9), 982–995.

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition*, *4*(3), 215–230.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *J. Exp. Psychol. Hum. Percept. Perform.*, *22*(4), 1005–1030.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.*, *18*, 643–662.

Tipper, S. P. (1985). The negative priming effect—Inhibitory priming by ignored objects. *Q. J. Exp. Psychol. [A]*, *37*(4), 571–590.

Tipper, S. P., & Behrmann, M. (1996). Object-centered not scene-based visual neglect. *J. Exp. Psychol. Hum. Percept. Perform.*, *22*(5), 1261–1278.

Treisman, A. (1969). Strategies and models of selective attention. *Psychol. Rev.*, *76*, 282–299.

Treisman, A. (1988). Features and objects—The 14th Bartlett Memorial Lecture. *Q. J. Exp. Psychol. [A]*, *40*(2), 201–237.

Treisman, A. (1993). The perception of features and objects. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control: A tribute to Donald Broadbent* (pp. 5–35). Oxford, UK: Clarendon Press.

Treisman, A. (1996). The binding problem. *Curr. Opin. Neurobiol.*, *6*(2), 171–178.

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cogn.*, *14*(4–8), 411–443.

Treisman, A., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 101–117). New York: Academic Press.

Treisman, A. M., & Gelade, G. (1980). Feature-integration theory of attention. *Cogn. Psych.*, *12*(1), 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision—Evidence from search asymmetries. *Psychol. Rev.*, *95*(1), 15–48.

Treisman, A. M., & Riley, J. G. A. (1969). Is selective attention selective perception or selective response—A further test. *J. Exp. Psychol.*, *79*, 27–34.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *J. Exp. Psychol. Hum. Percept. Perform.*, *16*(3), 459–478.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cogn. Psych.*, *14*(1), 107–141.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *J. Exp. Psychol. Hum. Percept. Perform.*, *24*(3), 830–846.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van Voorhis, S., & Hillyard, S. A. (1977). Visual evoked-potentials and selective attention to points in space. *Percept. Psychophys.*, *22*(1), 54–62.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends Cogn. Sci.*, *9*(12), 585–594.

Walsh, V., & Cowey, A. (1998). Magnetic stimulation studies of visual cognition. *Trends Cogn. Sci.*, *2*(3), 103–110.

Ward, L. M. (1982). Determinants of attention to local and global features of visual forms. *J. Exp. Psychol. Hum. Percept. Perform.*, *8*(4), 562–581.

Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychol. Rev.*, *104*(1), 90–122.

Welford, A. T. (1952). The psychological refractory period and the timing of high-speed performance: A review and a theory. *Br. J. Psychol.*, *43*, 219.

Wojciulik, E., & Kanwisher, N. (1999). The generality of parietal involvement in visual attention. *Neuron*, *23*(4), 747–764.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search—An alternative to the feature integration model for visual-search. *J. Exp. Psychol. Hum. Percept. Perform.*, *15*(3), 419–433.

Woodworth, R. S. (1938). *Experimental psychology*. New York: H. Holt and Company.

Wühr, P., & Frings, C. (2008). A case for inhibition: Visual attention suppresses the processing of irrelevant objects. *J. Exp. Psychol. Gen.*, *137*(1), 116–130.

Yamasaki, H., LaBar, K. S., & McCarthy, G. (2002). Dissociable prefrontal brain systems for attention and emotion. *Proc. Natl. Acad. Sci. USA*, *99*(17), 11447–11451.

Yantis, S., & Jonides, J. (1996). Attentional capture by abrupt onsets: New perceptual objects or visual masking? *J. Exp. Psychol. Hum. Percept. Perform.*, *22*(6), 1505–1513.

Yi, D.-J., Woodman, G. F., Widders, D., Marois, R., & Chun, M. M. (2004). Neural fate of ignored stimuli: Dissociable effects of perceptual and working memory load. *Nat. Neurosci.*, *7*(9), 992–996.

13 Mechanisms of Selective Attention in the Human Visual System: Evidence from Neuroimaging

SABINE KASTNER, STEPHANIE A. McMAINS, AND DIANE M. BECK

ABSTRACT In this chapter, we review evidence from functional brain imaging revealing that attention operates at various processing levels within the visual system including the lateral geniculate nucleus of the thalamus and the striate and extrastriate cortex. Attention modulates visual processing by enhancing neural responses to attended stimuli, attenuating responses to ignored stimuli, and increasing baseline activity in the absence of visual stimulation. These mechanisms operate dynamically on spatial locations, entire objects, or particular features, which constitute the units of selection. At intermediate cortical processing stages such as areas V4 and MT, the filtering of unwanted information is achieved by resolving competitive interactions among multiple simultaneously present stimuli. Together, these mechanisms allow us to select relevant information from the cluttered visual world in which we live to guide behavior.

Natural visual scenes are cluttered and contain many different objects. However, the capacity of the visual system to process information about multiple objects at any given moment in time is limited (e.g., Broadbent, 1958). Hence, attentional mechanisms are needed to select relevant information and to filter out irrelevant information from cluttered visual scenes. Selective visual attention is a broad term that refers to a variety of different behavioral phenomena. Directing attention to a spatial location has been shown to improve the accuracy and speed of subjects' responses to target stimuli that occur in that location (Posner, 1980). Attention also increases the perceptual sensitivity for the discrimination of target stimuli (Lu & Dosher, 1998), increases contrast sensitivity (Cameron, Tai, & Carrasco, 2002; Carrasco, Marie Giordano, & McElree, 2004), reduces the interference caused by distracters (Shiu & Pashler, 1995), and improves acuity (Carrasco, Loula, & Ho, 2006; Yeshurun & Carrasco, 1998).

SABINE KASTNER and STEPHANIE A. McMAINS Department of Psychology, Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey
DIANE M. BECK Department of Psychology, University of Illinois, Urbana-Champaign, Champaign, Illinois

The two most common behavioral paradigms employed to study visual attention are the spatial cuing paradigm that probes attention to a single location or stimulus (Posner, 1980) and the visual search task that probes attention in the presence of distracters (Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989). In the spatial cuing paradigm, subjects are instructed to maintain fixation and to direct attention covertly, that is, without shifting their gaze, to a peripheral target location, which is indicated by a cue. After a variable delay, a target stimulus, which subjects are required to detect, is presented briefly. On some trials, known as valid trials, the target appears at the cued (i.e., attended) location, and on other trials, known as invalid trials, the target appears at an uncued (i.e., unattended) location. The typically observed response difference in detecting stimuli on valid and invalid trials is thought to reflect the effects of attention on selected locations in space. In visual search tasks, subjects are given an array of stimuli (e.g., circles of different colors) and asked to report if a particular target stimulus (e.g., a red circle) is present in the array. Several factors affect performance in this task, such as the number of features that the target shares with other elements in the array. If the target (e.g., red circle) has a unique feature, such as being a different color from the distracters (e.g., green circles), the search is completed quickly, regardless of the number of elements in the array. This phenomenon is known as pop-out or efficient search. For other search arrays, where the target is defined by a conjunction of features (e.g., red horizontal line) that are shared by the distracters (e.g., red vertical and green horizontal lines), search time increases as a function of the number of elements in the array. This phenomenon is known as inefficient search, and the increase in search times is thought to reflect a serial search through the array for the target. However, under some circumstances, only a subset of the array needs to be searched. Simple features, such as color, can be used to *guide* search to just those elements that share a particular target feature (Wolfe et al., 1989). Visual search tasks have a clearer relationship than spatial cuing paradigms with our everyday

experience, where we typically face cluttered scenes with many objects that exceed our processing capacity. In this chapter, we will outline the neural basis that underlies attentional operations in these two tasks in the visual system, as they have been studied in humans using functional brain imaging and in nonhuman primates using electrophysiological techniques.

The units of selection

Attention is a highly flexible mechanism that can operate on regions of space, such as attending to a park on the right side of the road as you walk by, on particular features of an object, such as attending to the green leaves on a tree, or on entire objects, such as attending to the entire tree. Since the neural mechanisms mediating space-, object-, and feature-based attention have different characteristics, we will discuss each of them separately.

Space-Based Attention The spatial cuing paradigm is an example of a space-based selection process. In a typical fMRI study, the effects of space-based selection on neural responses have been investigated by presenting simple stimuli that activate the visual system well, such as flickering checkerboards to the left or right visual hemifield, while subjects directed attention to the stimulus (attended condition) or away from the stimulus (unattended condition) (e.g., O'Connor, Fukui, Pinsk, & Kastner, 2002). In the unattended condition, attention was directed away from the stimulus by having subjects count letters at fixation. The letter-counting task ensured proper fixation and effectively prevented subjects from covertly attending to the checkerboard stimuli. In the attended condition, subjects were instructed to covertly direct attention to the checkerboard stimulus and to detect luminance changes that occurred randomly in time at a peripheral stimulus location. Relative to the unattended condition, the mean fMRI signals evoked by a high-contrast checkerboard stimulus increased significantly in the attended condition in the lateral geniculate nucleus (LGN) and in visual cortex (figure 13.1*A*,*D*). In particular, attentional response enhancement was found in striate cortex, as well as in each extrastriate area along the ventral and dorsal pathways (figure 13.2*A*). Similar attentional response enhancement was obtained with activity evoked by a low-contrast checkerboard stimulus (figure 13.1*A*,*D*). Notably, these attention effects were shown to be spatially specific in other studies, in which identical stimuli were presented simultaneously to the right and left of fixation, while subjects were instructed to direct attention covertly to one or the other side (Brefczynski & DeYoe, 1999; Heinze et al., 1994; O'Connor et al.; Tootell et al., 1998). Taken together, these findings suggest that selective attention facilitates visual processing at thalamic and cortical stages by enhancing

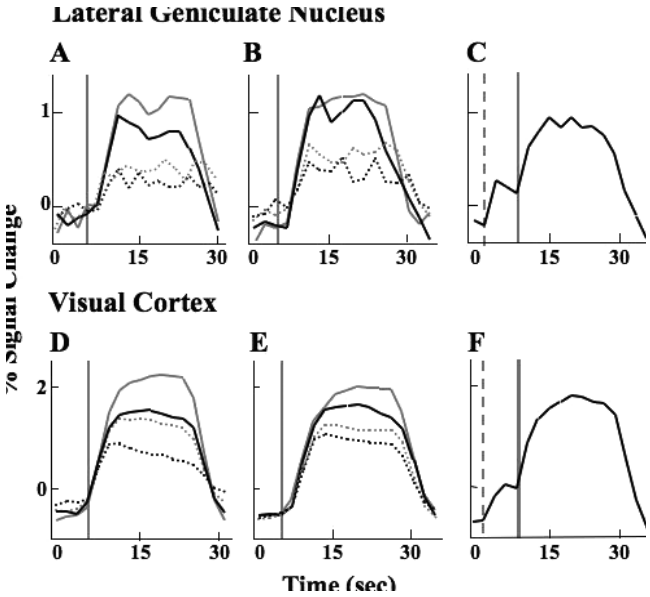


Figure 13.1 Attentional response enhancement, suppression, and increases in baseline activity in the LGN and in visual cortex. Group analysis ($n = 4$). Time series of fMRI signals in the LGN and visual cortex were combined across left and right hemispheres. Activity in visual cortex was pooled across areas V1, V2, V3/VP, V4, TEO, V3A, and MT/MST. (*A*, *D*) Attentional enhancement. During directed attention to the stimuli (gray curves), responses to both the high-contrast stimulus (100%, solid curves) and low-contrast stimulus (5%, dashed curves) were enhanced relative to an unattended condition (black curves). (*B*, *E*) Attentional suppression. During an attentionally demanding "hard" fixation task (black curves), responses evoked by both the high-contrast stimulus (100%, solid curves) and low-contrast stimulus (10%, dashed curves) were attenuated relative to an easy attention task at fixation (gray curves). (*C*, *F*) Baseline increases. Baseline activity was elevated during directed attention to the periphery of the visual hemifield in expectation of the stimulus onset; the beginning of the expectation period is indicated by the dashed vertical line. Gray vertical lines indicate the beginning of checkerboard presentation periods. (From O'Connor, Fukui, Pinsk, & Kastner, 2002.)

neural responses to an attended stimulus relative to those evoked by the same stimulus when ignored. Attentional response enhancement may be a neural correlate for behavioral attention effects such as increased accuracy and response speed or improved target discriminability (e.g., Posner, 1980; Lu & Dosher, 1998).

Spatial attention affects not only the processing of the selected information, but also the processing of the unattended information, which is typically the vast majority of incoming information. The neural fate of unattended stimuli was investigated in an fMRI experiment in which the attentional load of a task at fixation was varied (O'Connor et al., 2002). According to attentional load theory (Lavie & Tsal, 1994), the degree to which ignored stimuli are processed is determined by the amount of attentional capacity that is *not* dedicated to the selection process. This account predicts that

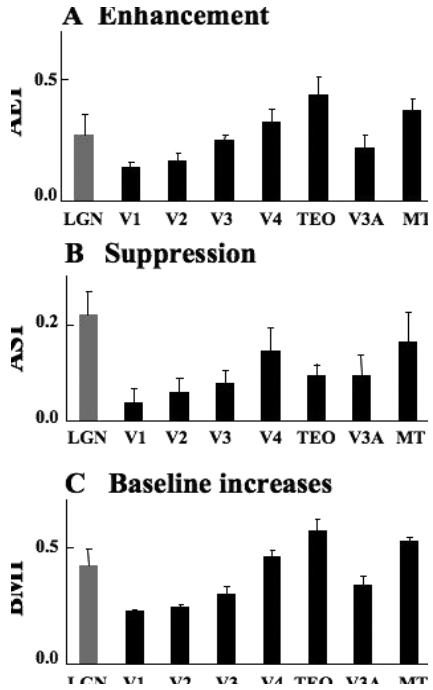


FIGURE 13.2 Attentional response modulation in the visual system. Attention effects that were obtained in the experiments presented in figure 13.1 were quantified by defining several indices: (*A*) attentional enhancement index (AEI), (*B*) attentional suppression index (ASI), (*C*) baseline modulation index (BMI). For all indices, larger values indicate larger effects of attention. Index values were computed for each subject based on normalized and averaged signals obtained in the different attention conditions and are presented as averaged index values from four subjects (for index definitions, see O'Connor et al., 2002). In visual cortex, attention effects increased from early to later processing stages. Attention effects in the LGN were larger than in V1. Vertical bars indicate standard error of the mean across subjects. (From O'Connor et al., 2002.)

neural responses to unattended stimuli should be attenuated depending on the attentional load necessary to process the attended stimulus. This idea was tested by using the checkerboard paradigm described earlier while subjects performed either an easy (low-load) attention task or a hard (high-load) attention task at fixation and ignored the peripheral checkerboard stimuli. Relative to the easy-task condition, mean fMRI signals evoked by the high-contrast and by the low-contrast stimuli decreased significantly in the hard-task condition across the visual system with the smallest effects in early visual cortex and the largest effects in LGN and extrastriate cortex (figure 13.1*B*,*E*). Taken together, these findings suggest that neural activity evoked by ignored stimuli is attenuated at several stages of visual processing as a function of the load of attentional resources engaged elsewhere (O'Connor et al., 2002; Rees, Frith, & Lavie, 1997; Schwartz et al., 2005). Attentional-load-dependent suppression of unattended stimuli may be a neural correlate for behavioral effects such as reduction of interference caused by distracters (Shiu & Pashler, 1995). Further, there is evidence that the enhancement of activity at an attended location and the suppression of activity at unattended locations operate in a push-pull fashion and thus represent codependent mechanisms (Pinsk, Doniger, & Kastner, 2004; Schwartz et al., 2005).

An important component of the Posner task is the cuing period during which subjects deploy attention to a location in space at which visual stimuli are expected to occur. A neural correlate of cue-related activity has been found in physiology studies demonstrating that spontaneous (baseline) firing rates were 30–40% higher for neurons in areas V2 and V4 when the animal was cued to attend covertly to a location within the neuron's receptive field (RF) before the stimulus was presented there—that is, in the absence of visual stimulation (Lee, Williford, & Maunsell, 2007; Luck, Chelazzi, Hillyard, & Desimone, 1997; but see McAdams & Maunsell, 1999). This increased baseline activity has been interpreted as a direct demonstration of a top-down signal that feeds back from higher-order to lower-order areas. In the latter areas, this feedback signal appears to bias neurons representing the attended location, thereby favoring stimuli that will appear there at the expense of those appearing at unattended locations.

To investigate attention-related baseline increases in the human visual system in the absence of visual stimulation, fMRI activity was measured while subjects were cued to covertly direct attention to the periphery of the left or right visual hemifield and to expect the onset of a stimulus (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; O'Connor et al., 2002; Sylvester, Shulman, Jack, & Corbetta, 2007). The expectation period, during which subjects were attending to the periphery without receiving visual input, was followed by attended presentations of a high-contrast checkerboard. During the attended presentations, subjects counted the occurrences of luminance changes. Relative to the preceding blank period in which subjects maintained fixation at the center of the screen and did not attend to the periphery, fMRI signals increased during the expectation period in the LGN and the striate and extrastriate cortex (figures 13.1*C*,*F*, 13.2*C*). This elevation of baseline activity was followed by a further response increase evoked by the visual stimuli (figure 13.1*C*).

Similar to response modulation, the magnitude of increases in baseline activity depends on several variables, including the expected task difficulty (Ress, Backus, & Heeger, 2000) or the expected presence or absence of distracter stimuli (Serences, Yantis, Culberson, & Awh, 2004). Early studies have found evidence that baseline increases are feature specific; that is, they are stronger during the expectation of a preferred compared to a nonpreferred stimulus feature in areas that preferentially process a particular stimulus feature (e.g., color in area V4 or motion in area MT) (Chawla, Rees, & Friston, 1999; Shulman et al., 1999). However, more

recent studies have not confirmed these initial findings (McMains, Fehd, Emmanouli, & Kastner, 2007), and therefore the feature specificity of baseline increases is still an open question. The baseline increases found in human visual cortex may be subserved by increases in spontaneous firing rate similar to those found in the single-cell recording studies (Luck et al., 1997), but summed over large populations of neurons. The increases evoked by directing attention to a target location in anticipation of a behaviorally relevant stimulus at that attended location are thus likely to reflect a top-down feedback bias in favor of a preferred stimulus at an attended location.

Comparison of space-based attention effects across the visual system With fMRI, neural responses can be investigated at the population level and across a wide range of different processing stages, allowing for quantitative comparisons of attentional modulatory effects across the visual system. For example, the attention effects of enhancement, suppression, and baseline increases described in the last section can be quantified in each visual area by calculating an index value. Large index values indicate large effects of attention (for further details, see O'Connor et al., 2002). As shown in figure 13.2, the magnitude of all attention effects increased from early to more advanced processing stages along both the ventral and dorsal pathways of visual cortex (figure 13.2*A–C*; Cook & Maunsell, 2002; Kastner, De Weerd, Desimone, & Ungerleider, 1998; Martinez et al., 1999; Mehta, Ulbert, & Schroeder, 2000). This finding is consistent with the idea that attention operates through top-down signals that are transmitted via corticocortical feedback connections in a hierarchical fashion. Thereby, areas at advanced levels of visual cortical processing are more strongly controlled by attentional mechanisms than are early processing levels. This idea is supported by single-cell recording studies, which have shown that attentional effects in area TE of inferior temporal cortex have a latency of approximately 150 ms (Chelazzi, Duncan, Miller, & Desimone, 1998), whereas attentional effects in V1 have a longer latency, approximately 230 ms (Roelfsema, Lamme, & Spekreijse, 1998). According to this account, one would predict smaller attention effects in the LGN than in striate cortex. Surprisingly, it was found that all attention effects tended to be larger in the LGN than in striate cortex (figure 13.2*A–C*). This finding raises the possibility that attentional modulation in the LGN may not be exclusively attributable to corticothalamic feedback from striate cortex, but may also reflect additional modulatory influences from other sources. In addition to corticothalamic feedback projections from V1, the LGN receives inputs from the superior colliculus, which is part of a distributed network of areas controlling eye movements, and the thalamic reticular nucleus (TRN), which has long been implicated in theoretical accounts of selective attention (Crick, 1984; Sherman & Guillery, 2001).

Object-Based Attention In addition to selecting regions of space, attention can also be directed to entire objects. Object-based attention signals have been investigated in human neuroimaging studies, in which subjects were asked to attend to a particular feature of an object, while signals evoked by the unattended features of the attended object were measured (McMains et al., 2007; O'Craven, Downing, & Kanwisher, 1999). In one such study, Kanwisher and colleagues used stimuli that consisted of overlapping images of houses and faces (O'Craven et al., 1999; figure 13.3*A*). On any given trial, one of those images moved. Subjects attended to and performed a task on either the house, the face, or the motion, resulting in increased activity in regions specialized for processing the attended feature or object, that is, the fusiform face area, parahippocampal place area, or MT, respectively. Interestingly, activity also spread to the unattended feature of the attended stimulus. For instance, when the face stimuli were moving and subjects attended to the faces, increased activity was observed in MT as compared to when subjects attended to stationary house stimuli (figure 13.3*B*). Thus attentional biasing signals appear to spread to the unattended features of the attended object, resulting in enhanced signals in regions specialized for processing the unattended feature. These studies demonstrate that processing can be biased in favor of an attended object, with all features of the attended object receiving some amount of enhanced processing. In addition, when attention is directed to only a portion of an object, attention has been found to spread throughout the entire object (Muller & Kleinschmidt, 2003; Roelfsema et al., 1998). In a physiology study by Roelfsema and colleagues, monkeys were presented with two curved lines, one of which was the target curve. Small changes in the stimulus close to fixation determined which line was the target, while neuronal responses were measured in V1 neurons with RFs centered over distant points on one of the lines. Responses were increased when the line segment in the RF was part of the target curve compared to when it was part of the distractor curve. Together, these results suggest that when attention is directed to an object, all features of the attended object are enhanced along with the entire spatial extent of the attended object. These findings may provide a neural correlate for classical behavioral studies showing that subjects perform worse at reporting two attributes that belong to different objects as compared to two attributes belonging to the same object (Duncan, 1984).

Feature-Based Selection In addition to biasing processing in favor of a spatial location or an entire object,

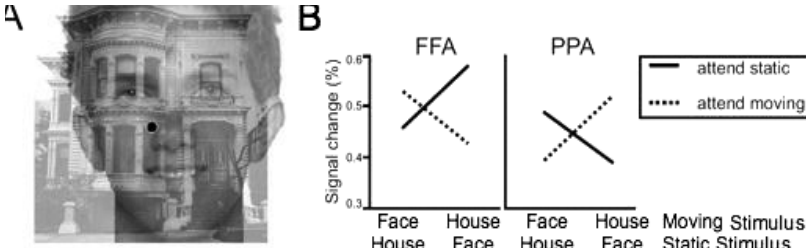


FIGURE 13.3 Object-based attention in visual cortex. (*A*) Example stimulus from experiment 2 of the study by O'Craven, Downing, and Kanwisher (1999) demonstrating object-based attention. The stimuli consisted of overlapping house and face stimuli. On each trial either the house or the face stimulus moved while subjects performed a consecutive repetition-detection task on either the direction of motion of the moving stimuli or the position of the stationary object, which was offset slightly from trial to trial. (*B*) Averaged fMRI signals ($n = 4$), computed as percent signal change, for the fusiform face area (FFA) and the parahippocampal place area (PPA). Solid lines represent activity when subjects attended to the static stimulus, and dotted lines represent trials during which subjects attended to the moving stimulus. Activity was higher in the FFA when "faceness" was the irrelevant property of the attended object (Attend Moving, Face Moving) than when it was a property of the unattended object (Attend Moving, House Moving). The response pattern of the PPA was identical for "houseness." (Modified figure 42.1, Freiwald & Kanwisher, 2004.)

attention can also bias processing in favor of a particular stimulus attribute, or feature. In experiments investigating feature-based attention, stimuli are typically composed of multiple stimulus features (e.g., colored shapes moving in different directions), and subjects are cued to attend to a particular feature dimension (e.g., the color red) while ignoring the other dimensions. In monkey physiology studies, neural responses increased when the feature in the RF matched the cued feature regardless of where the animal was attending. Feature-based attention effects have been observed in area V4 for several feature dimensions (for a review see Maunsell & Treue, 2006) including color (Motter, 1994), luminance (Motter, 1994), and orientation (Haenny, Maunsell, & Schiller, 1988), and in MT for direction of motion (Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999). In addition, the temporal characteristics of feature-based and space-based attention were investigated in area V4 (Hayden & Gallant, 2005). Feature-based attention effects were found to be sustained throughout the visual evoked responses, whereas space-based attention effects were more transient, peaking in the later portion of the response. These findings suggest that space- and feature-based attention effects rely on different neural mechanisms.

In human neuroimaging studies, where activity is measured at the level of entire areas within neural networks, researchers have taken advantage of the functional specialization of visual cortex in studying feature-based attention (Beauchamp, Cox, & DeYoe, 1997; Buechel et al., 1998; Clark et al., 1997; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991; McMains et al., 2007; O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997; Serences & Boynton, 2007; Sohn, Chong, Papathomas, & Vidnyanszky, 2005). In one such study, Corbetta and colleagues investigated attention to shape, color, or speed and observed enhanced activity in visual regions specialized for processing the attended feature, including enhanced activity in the posterior fusiform gyrus for attention to shape, area V4 for attention to color, and area MT for attention to motion.

Feature-based attention mechanisms have also been investigated in the presence of distracters. The observation that neural responses to a selected feature increase regardless of where the animal attends has led to the hypothesis that feature-based attention may operate globally throughout the visual field (Bichot, Rossi, & Desimone, 2005; Martinez-Trujillo & Treue, 2004; McAdams & Maunsell, 2000; Saenz, Buracas, & Boynton, 2002; Serences & Boynton, 2007). If one considers a visual search task where subjects are looking for a red circle in an array of colored shapes, it will certainly be advantageous from a computational point of view to increase neural responses to any red items, thereby marking the candidate target stimuli and restricting the remaining search to the subset of red shapes to ultimately find the circle. This approach is opposed to space- and object-based attention, which are both inherently tied to a spatial location. This hypothesis has been tested in a physiology study where monkeys performed a visual search task (Bichot et al., 2005). Neuronal responses were enhanced when the stimulus inside the RF was the same color or shape as the target stimulus. This result occurred throughout the search period, regardless of where the monkey was attending.

A similar effect has been observed in humans in an experimental design with two stimuli, one presented in each visual hemifield (Saenz et al., 2002). The attended stimulus consisted of two overlapping dot patterns, one moving upward and the other moving downward. The stimulus in the unattended hemifield always moved in the same direction (e.g., downward). Functional MRI signals were measured in the retinotopic representation of the unattended stimulus while subjects alternated between attending to the same direction as the distracter (i.e., downward) or to the opposite direction

(i.e., upward). As in the physiology studies, signals evoked by the distracter stimulus were greater in V1, V2, V3, V3A, V4, and MT when it matched the attended direction. In addition, it has been shown that enhanced responses to the attended feature spread to regions of the visual field that are not stimulated (Serences & Boynton, 2007). Taken together, these results suggest a mechanism for feature-based attention that operates globally and biases processing throughout the visual field by enhancing the population response to attended features. This type of global attention mechanism might be important in guiding spatial attention to regions of the visual field that contain behaviorally relevant stimuli and may be one of the mechanisms that underlie guided visual search (Wolfe et al., 1989).

Summary Mechanisms for the selection of behaviorally relevant information operate flexibly in the visual system on spatial locations, features of objects, or entire objects by enhancing responses evoked by attended stimuli, by suppressing responses evoked by unattended stimuli, and by increasing neural baseline activity to facilitate the processing of expected stimuli. These different effects of attention provide a framework for the neural basis of simple attentional operations such as orienting to a spatial location, as tested in the spatial cuing paradigm. In the next section, we will explore the effects of attention when multiple stimuli are present, such as in visual search tasks.

Selection among multiple competing objects

One of the key aspects of the visual search paradigm, and indeed what makes it relevant to our everyday behavior, is that a target must be found among a field of other objects. The defining role of selective attention in this case is to mediate the selection of a subset of the available information for further processing. As such, in order to understand selective attention, it is necessary to understand first how the brain processes multiple stimuli simultaneously present in the visual field.

Neural Basis of Competition A large body of evidence from both single-cell physiology and neuroimaging suggests that multiple stimuli present at the same time within a neuron's RF are not processed independently, but interact with each other in a mutually suppressive way. In physiology studies (Miller, Gochin, & Gross, 1993; Recanzone, Wurtz, & Schwarz, 1997; Reynolds, Chelazzi, & Desimone, 1999; Snowden, Treue, Erickson, & Andersen, 1991), the responses to paired stimuli were found to be smaller than the sum of the responses evoked by each stimulus individually. In particular, the response evoked by a pair of stimuli was a weighted average of the individual responses (Luck et al., 1997; Reynolds et al.). Suppressive interactions such as these among multiple stimuli have been found in several visual areas in the monkey brain, including V2, V4, MT, MST, and IT (Miller et al.; Recanzone et al.; Reynolds et al.; Snowden et al.), and have been interpreted as the neural substrate of an ongoing competition among multiple simultaneously present stimuli for representation in visual cortex.

In the human brain, evidence for neural competition has been found using an fMRI paradigm (Beck & Kastner, 2005, 2007; Kastner et al., 2001), in which four colorful and patterned visual stimuli that optimally activate ventral visual cortex were presented in four nearby locations to the periphery of the visual field, while subjects maintained fixation (figure 13.4*A*). Critically, these stimuli were presented under two different presentation conditions, sequential and simultaneous. In the sequential presentation condition, each stimulus was presented alone in one of the four locations, one after the other. In the simultaneous presentation condition, the same four stimuli appeared simultaneously in the four locations. Thus, integrated over time, the physical stimulation parameters were identical in each of the four locations in the two presentation conditions. However, competitive interactions among stimuli could take place only in the simultaneous and not in the sequential presentation condition.

Consistent with the physiology literature, simultaneous presentations evoked weaker responses than sequential presentations in areas V1, V2/VP, V4, TEO, V3A, and MT. The response differences were smallest in V1 and increased in magnitude toward ventral extrastriate areas V4 (figure 13.4*D*) and TEO, and dorsal extrastriate areas V3A and MT. In other words, the suppressive interactions among the four stimuli scaled with the increasing RF sizes across visual cortex, consistent with the idea that the stimuli were competing for representation at the level of the RF.

The idea that suppressive interactions are scaled to RF size was tested directly by varying the spatial separation among the stimuli (Kastner et al., 2001). If stimuli are competing at the level of the RF, then increasing the spatial separation among stimuli should decrease the level of competition within the RF, and decreasing the array size should increase competitive interactions in areas with smaller RFs. In keeping with both of these predictions, suppressive interactions were twice as strong in V1 and V2 with a $2 \times 2°$ display as compared to those induced with a $4 \times 4°$ display, and separating the stimuli in the $2 \times 2°$ display by 4° abolished suppressive interactions in V2 and reduced them in V4, but did not affect them in TEO. Separating the stimuli by 6° led to a further reduction of suppression effects in V4, but again had no effect in TEO. Taken together, monkey physiology and human brain imaging studies have begun to establish a neural basis for competitive representations of stimuli in the visual system.

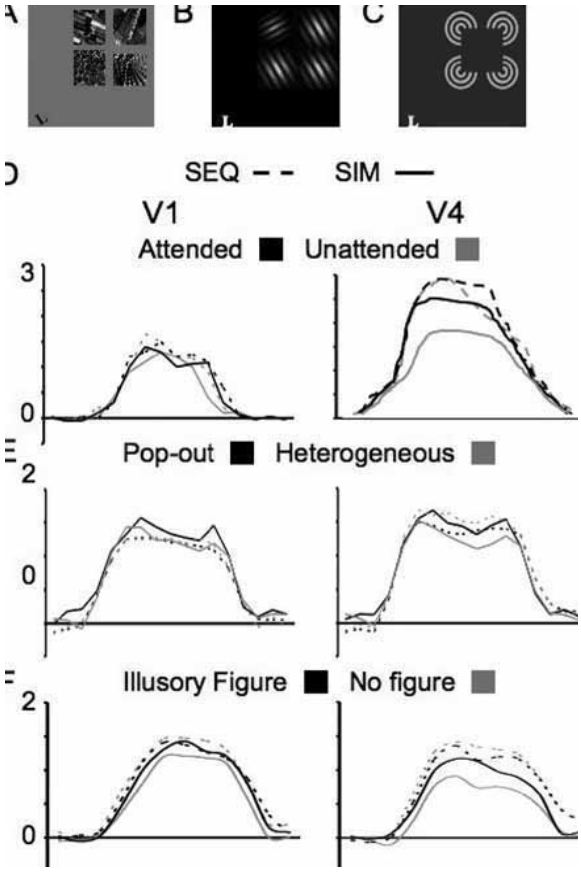


FIGURE 13.4 Top-down and bottom-up biases in resolving neural competition in visual cortex. In three experiments, competitive interactions among stimuli were assessed in the presence of top-down and bottom-up biases. Competition was measured by comparing fMRI responses to four stimuli presented either simultaneously (potentially competing; SIM) or sequentially (noncompeting; SEQ). (*A*) In the study investigating top-down attention biases (Kastner, De Weerd, Desimone, & Ungerleider, 1998), the four stimuli were complex colored images. In the attended conditions, subjects monitored the lower left location for the appearance of a target stimulus, whereas in the unattended condition they ignored the colorful stimuli and instead detected targets at fixation. (*B*) In the bottom-up pop-out study (Beck & Kastner, 2005), the stimuli were four Gabor patches that either all differed in color and orientation (heterogeneous condition; not shown) or in which one Gabor differed in color and orientation from the rest (pop-out condition; shown here without color). (*C*) In the illusory contour study (McMains & Kastner, 2007), four "Pacman" images were aligned to form an illusory square (shown) or rotated such that no square was perceived. (*D–F*) Dashed curves indicate activity evoked by sequential presentations and solid curves indicate activity evoked by simultaneous presentations for attention (*D*), pop-out (*E*), and illusory contour (*F*) studies. Black lines represent the conditions in which a top-down or bottom-up bias was probed, and gray lines represent conditions during which stimuli were competing without any bias. For all three studies, competition was partially overcome in V4 when a bias was present (black curves), resulting in smaller differences in responses evoked by sequential and simultaneous presentations (dashed and solid curves). These effects were smaller (*F*) or nonexistent in V1 (*D*, *E*).

It is important to note that the suppressive (competitive) interactions across visual cortex discussed thus far occurred automatically and in the absence of attentional allocation to the stimuli. In fact, in these paradigms, participants were engaged in an attention-demanding task at fixation. Thus neural competition would appear to be pervasive in the representation of cluttered visual scenes. In order to overcome this less than optimal representation of objects within a visual scene, there need to be mechanisms by which this ongoing competition among multiple stimuli can be resolved. The allocation of top-down attention is one such mechanism.

TOP-DOWN INFLUENCES ON COMPETITION Single-cell recording studies have shown that spatially directed attention represents one top-down mechanism that can bias the competition among multiple stimuli in favor of the attended stimulus by modulating competitive interactions. When a monkey directed attention to one of two competing stimuli within an RF, the responses in extrastriate areas V2, V4, and MT to the pair of stimuli were as large as those to that stimulus presented alone (Recanzone & Wurtz, 2000; Reynolds et al., 1999). A similar mechanism appears to operate in the human visual cortex. Kastner and colleagues (1998) studied the effects of spatially directed attention on multiple competing visual stimuli in a variation of the paradigm described earlier. In addition to the two different presentation conditions, sequential and simultaneous, two different attentional conditions were tested, attended and unattended. During the unattended condition, attention was directed away from the peripheral visual display by having subjects count letters at fixation. In the attended condition, subjects were instructed to attend covertly to the peripheral stimulus location in the display closest to fixation and to count the occurrences of a target stimulus. Directing attention to this location led to greater activity increases for simultaneously presented stimuli than for sequentially presented stimuli in areas V4 and TEO, but not in early visual areas such as V1 (figure 13.4*D*). Like the competition effects, the magnitude of the attentional effects scaled with RF size, with the strongest reduction of suppression occurring in ventral extrastriate areas V4 and TEO (see also Bles, Schwarzbach, De Weerd, Goebel, & Jansma, 2006). It is important to note that the stage at which these attention mechanisms will operate in the visual system is flexibly determined by the spatial scale of the displays, similar to the spatial scaling of competition mechanisms to RF size. For example, targets in displays of large spatial scale will engage areas with larger RFs and will shift the locus of attentional selection to more anterior extrastriate areas relative to targets in displays of smaller spatial scale (Buffalo, Bertini, Ungerleider, & Desimone, 2005; Hopf et al., 2006). Together, these findings support the idea that directed attention

enhances information processing of stimuli at the attended location by counteracting suppression induced by nearby stimuli. This may be an important mechanism by which unwanted information is filtered out from nearby distracters.

Bottom-Up Influences on Competition Related to Visual Salience Top-down attention, however, is not the only mechanism that can resolve competition in visual cortex. Properties of the stimulus—that is, bottom-up stimulus-driven mechanisms—have also been shown to modulate competition (Beck & Kastner, 2005, 2007; McMains & Kastner, 2007; Reynolds & Desimone, 2003). For example, Beck and Kastner (2005) found that competition could be influenced by visual salience, a bottom-up stimulus-driven mechanism that does not depend on the focus of attention. Instead of the complex images used in the previous studies, in which the sequential/simultaneous paradigm was investigated, they used four Gabor patches of different colors and orientations. These Gabor patches could be presented in two display contexts: pop-out displays, in which a single item differed from the others in color and orientation (figure 13.4*B*), and heterogeneous displays, in which all items differed from each other in both dimensions. As has been shown with the heterogeneous displays used in previous experiments, the heterogeneous displays produced robust suppressive interactions among multiple stimuli in areas V2/VP and V4 (figure 13.4*E*). However, this suppression was eliminated when the same stimuli were presented in the context of pop-out displays, consistent with the prediction that visual salience can bias competitive interactions among multiple stimuli in intermediate processing areas (figure 13.4*E*).

As in previous studies, no evidence of competitive interactions was found in area V1, presumably because of the small RF sizes in that area. However, an effect of display context was evident in this early visual area: simultaneously presented pop-out displays evoked more activity than any of the other three conditions, consistent with both physiology studies and computational models suggesting that pop-out may be computed as early as in area V1 (Kastner, Nothdurft, & Pigarev, 1999; Knierim & Van Essen, 1992; Li, 1999; Nothdurft, Gallant, & Van Essen, 1999). Taken together, these data suggest that V1 may be the source of the signal that biases neural competition in extrastriate cortex when salient stimuli are present in the visual scene, a finding which further distinguishes this bottom-up bias from top-down biases that are thought to have their source in frontoparietal cortex (Kastner et al.; T. Moore & Armstrong, 2003).

Bottom-Up Influences on Competition Related to Scene Segmentation Both top-down attention and bottom-up visual salience constitute a spatial bias that results in a single item dominating the response of the neuron. However, as described in the section "The units of selection," attention not only selects a location or feature, but can also select whole objects, including all the features that comprise the object. The existence of object-based selection argues against competition among the features within an object and instead argues for mechanisms that bind the features of an object together across multiple areas prior to competition among objects. Indeed, there are a number of factors that are thought to influence how the components of a scene are structured into larger units such as objects and groups of objects. The first of these *principles of perceptual organization* were proposed by the Gestalt psychologists (e.g., Wertheimer, 1923; Rubin, 1958). For instance, visual stimuli may be perceptually grouped according to their similarity, proximity, common fate (Wertheimer, 1923), and other stimulus properties (Palmer & Rock, 1994; Palmer, 1992), thereby linking elements of a scene that are likely to belong together, segmenting the scene into a more limited number of objectlike perceptual units. There is growing evidence that some forms of perceptual organization do not require top-down control, but rather represent automatic bottom-up processes (Altmann, Bulthoff, & Kourtzi, 2003; Driver, Baylis, & Rafal, 1992; Duncan, 1984; Kastner, De Weerd, & Ungerleider, 2000; Lamy, Segal, & Ruderman, 2006; Marcus & Van Essen, 2002; C. Moore & Egeth, 1997; Qiu, Sugihara, & von der Heydt, 2007; Russell & Driver, 2005; but see also Ben-Av, Sagi, & Braun, 1992; Han, Jiang, Mao, Humphreys, & Gu, 2005; Mack, Tang, Tuma, Kahn, & Rock, 1992; Roelfsema, 2006). Therefore, perceptual organization principles may represent bottom-up stimulus-driven processes by which features within an object are grouped, allowing competition to occur among objects and not among features within an object (Desimone & Duncan, 1995). According to this view, any principle of perceptual organization that produces a bottom-up bias may modulate competition.

Evidence in favor of this prediction was found in a series of studies that used variants of the sequential/simultaneous fMRI paradigm to investigate effects of perceptual organization on competition. In one study, suppressive interactions among four identical items (homogeneous display) were compared to those induced by four stimuli that differed in both color and orientation (heterogeneous display). Because identical or similar items that are present in nearby locations tend to form perceptual groups by the Gestalt principle of similarity, the prediction was tested that competitive interactions should be minimal with identical stimuli in the display (homogeneous condition) as compared to the heterogeneous condition. In accordance with previous data, simultaneous presentation of four heterogeneous visual stimuli evoked significantly less activity in areas V2, VP, and V4 than the same stimuli presented sequentially. However, when the four

stimuli were identical, the suppression was considerably reduced relative to the heterogeneous conditions. This result suggests that grouping by similarity represents a bottom-up bias that may influence or even determine the amount of competition among items.

Although grouping is arguably the most well-known of the perceptual organization processes, there are a number of other processes critical to our ability to segment and organize a scene. Before objects can be grouped together, the visual system must decide what regions in the scene constitute potential objects; that is, it must segment figure from ground (Rubin, 1958). Further complicating this process is the fact that some potential objects may be partially occluded from view, requiring the visual system to infer the presence of objects on the basis of the information present in the scene; that is, it must rely on visual interpolation mechanisms (Palmer, 1999). Both of these processes were probed in a second study (McMains & Kastner, 2007), using the Kanizsa illusion (Kanizsa, 1976). In the Kanizsa illusion, four circular "Pacman" items, also called inducers, are aligned to form an illusory square (figure 13.4*C*) that is perceived as a single foreground element with the inducers lying behind it. When the four inducers are rotated inward, the illusion occurs as a result of the assignment of the L-shaped borders to a common object, but does not occur when they are rotated outward. Based on the hypothesis that the degree of perceptual organization in a visual display should determine the degree of competition, it was predicted that when the four inducers were rotated inward giving rise to the illusion and were thus part of a single object, they should not compete with each other. Alternatively, if the four inducers were rotated outward, thereby disrupting the illusion, they would be treated as four separate objects, which compete for neural representation independently. As predicted, the competitive interactions were significantly reduced across visual cortex when the four inducers formed a single foreground object, but not when they were rotated outward representing four separate items (figure 13.4*F*). Thus far, a number of perceptual organization processes, such as grouping by the Gestalt factor of similarity and figure-ground segmentation and visual interpolation processes related to illusory contour formation, have been shown to affect the outcome of competitive interactions, thereby suggesting that any form of perceptual organization might influence the ongoing competition (Beck & Kastner, 2007; McMains & Kastner, 2007).

The relationship between perceptual organization principles and competition can be interpreted in at least two ways. Competition may be influenced by mechanisms that mediate perceptual organization from elsewhere in the cortex. These mechanisms could boost the activity related to the set of stimuli as it enters V4, effectively counteracting any competition that may have occurred between stimuli. Such a perspective is consistent with effects of Gestalt grouping and figure-ground segmentation found in early visual cortex (Kapadia, Ito, Gilbert, & Westheimer, 1995; Kastner et al., 2000; Lamme, 1995; Nothdurft et al., 1999; Qiu et al., 2007; Zhou, Friedman, & von der Heydt, 2000). Alternatively, the degree to which perceptual organization occurs may be a consequence of competitive interactions. As mentioned, the response of V4 neurons to a pair of stimuli is best described as a weighted average of the responses to the two stimuli when presented alone (Luck et al., 1997; Reynolds et al., 2000). If the two stimuli that comprise the pair are identical, as in the grouping-by-similarity study, then the weighted-average model would predict that the response to the pair should be indistinguishable from the response to each of the individual stimuli (Reynolds et al.). Thus there may not be any need to appeal to additional grouping mechanisms to explain these findings. Instead, the reduced competition present in the displays with identical items, relative to the one with different stimuli, may simply be the result of the averaging procedure performed by the neurons in areas such as V4. If less competition is evoked by items that are perceptually organized, then there is no need to select or filter any one of them, and instead the items are processed as a group. Importantly, these are not mutually exclusive possibilities. Further, it is unlikely that these interactions can be explained by a unified neural mechanism. Rather, the variety of perceptual organization principles may rely on a variety of underlying neural processes.

Relation of Bottom-Up and Top-Down Mechanisms

The studies described thus far suggest that both top-down and bottom-up processes can bias competitive interactions in visual cortex. How might these processes interact? Evidence comes from a physiology study (Qiu et al., 2007) in which the effects of attention on image segmentation processes were probed in V2 neurons. Neurons in V2 have previously been found to integrate contextual information from beyond their small RFs in order to signal when a border in their RF belongs to an attended object, an effect termed border ownership (Zhou et al., 2000). Consistent with previous studies (Driver et al., 1992; Kastner et al., 2000; Lamy, Segal, & Ruderman, 2006; C. Moore & Egeth, 1997), V2 neurons signaled border ownership for both attended and unattended figures, suggesting that attention is not necessary for figure-ground segmentation (Qiu et al.). When the interaction of attention and border ownership was investigated, the magnitudes of the attention effects were found to be predicted by the neurons' border ownership responses, such that attention effects were larger when the object that owned the border in the RF was attended to compared to when a different figure of equal distance away from the RF was attended to (figure 13.5*E*). These results were interpreted using a novel framework, the "interface

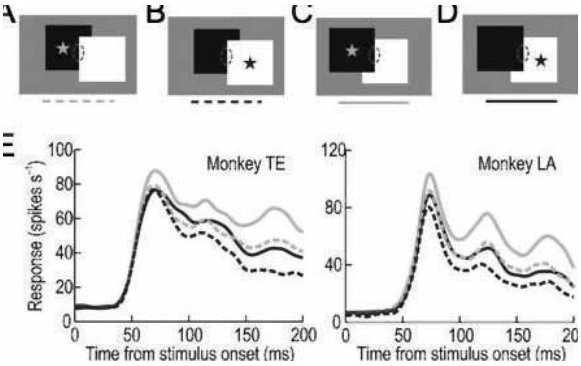


FIGURE 13.5 Responses of neurons influenced by border ownership and attention. (*A–D*) Example stimuli for the different experimental conditions. Asterisks indicate the location of attention, and the ellipses symbolize the RF of the neuron that was recorded from. (*A–B*) The right figure was in front, with the nonpreferred side of the object in the RF. (*C–D*) The left figure was in front, with the preferred side of the object in the RF. (*E*) Firing rates as a function of time for the population of neurons influenced by border ownership and attention for each of two monkeys. The solid lines represent preferred border ownership, and dashed lines indicate nonpreferred border ownership; gray indicates attention to the left side, and black indicates attention to the right side. Attention effects were always greater when the animals attended to the left, that is, to the preferred object side ($A > B$ and $C > D$). However, attention effects were largest when the animals attended to the preferred object side (left) and the border in the RF was "owned" by the object on the left (*C*). (Modified from figure 5 of Qiu, Sugihara, & von der Heydt, 2007.)

hypothesis of attention." This hypothesis suggests that the correlation between border ownership responses and attentional effects may be explained by assuming that the same circuits underlie both processes and that the circuit that mediates border ownership provides an interface for attentional mechanisms. Thus top-down mechanisms might operate on intrinsic local circuits within V2 or on a local V2–V4 feedback network that mediates border ownership during the selection of object information.

The effects of attention on local circuits related to figure-ground segmentation in early visual cortex may represent one example in support of the interface hypothesis. A similar account may explain selection processes operating on neural competition at intermediate processing stages such as V4. As described earlier, when a stimulus is selected for further processing among multiple items, top-down processes appear to counteract the suppressive influences induced by nearby stimuli, thereby strengthening the neural representation of the attended stimulus. It is possible that the local intrinsic circuit engaged during competition within these areas provides an interface for the selection mechanisms to operate on. Consistent with this idea, local interneurons in V4, which are thought to subserve local circuits, receive stronger attentional modulation than other cell classes (Mitchell, Sundberg, & Reynolds, 2007). Thus the interface hypothesis may be a useful novel framework for the interpretation of a number of empirical findings in the attention field.

SUMMARY One of the primary roles of attention is to select a small set of items among many, as when searching for an object in a cluttered visual scene. The need for such a mechanism becomes clearer when we consider the mutually suppressive interactions among multiple stimuli in visual cortex. Directed attention has been shown to bias this competition in favor of an attended stimulus, thus restoring a more optimal representation of that stimulus. Similarly, competitive interactions can be modulated by bottom-up factors such as salience and factors related to perceptual organization. Although these bottom-up factors represent qualitatively different biases from top-down directed attention, it also appears that the two factors interact to organize and ultimately select relevant aspects of a visual scene.

Conclusion

Evidence from functional brain imaging reveals that attention operates at various processing levels within the human visual system and beyond. These attention mechanisms appear to be controlled by a distributed network of higher-order areas in frontal and parietal cortex, which generate top-down signals that are transmitted by way of feedback connections to the visual system (see chapter 14 by Corbetta, Sylvester, & Shulman). Together, these widely distributed brain systems cooperate to mediate the selection of behaviorally relevant information that can be further utilized in other cognitive networks to ultimately guide goal-directed action.

REFERENCES

ALTMANN, C. F., BULTHOFF, H. H., & KOURTZI, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Curr. Biol.*, *13*(4), 342–349.

BEAUCHAMP, M. S., COX, R. W., & DEYOE, E. A. (1997). Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *J. Neurophysiol.*, *78*, 516–520.

BECK, D. M., & KASTNER, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nat. Neurosci.*, *8*, 1110–1116.

BECK, D. M., & KASTNER, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *J. Vis.*, *7*(2), 19.11–19.12.

BEN-AV, M. B., SAGI, D., & BRAUN, J. (1992). Visual attention in perceptual grouping. *Percept. Psychophys.*, *52*(3), 277–294.

BICHOT, N. P., ROSSI, A. F., & DESIMONE, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*(5721), 529–534.

BLES, M., SCHWARZBACH, J., DE WEERD, P., GOEBEL, R., & JANSMA, B. M. (2006). Receptive field size-dependent attention

effects in simultaneously presented stimulus displays. *NeuroImage*, *30*(2), 506–511.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the "spotlight" of visual attention. *Nat. Neurosci.*, *2*(4), 370–374.

Broadbent, D. E. (1958). *Perception and communication*. Oxford, UK: Oxford University Press.

Buechel, C., Josephs, O., Rees, G., Turner, R., Frith, C. D., & Friston, K. J. (1998). The functional anatomy of attention to visual motion: A functional MRI study. *Brain*, *121*(Pt. 7), 1281–1294.

Buffalo, E. A., Bertini, G., Ungerleider, L. G., & Desimone, R. (2005). Impaired filtering of distracter stimuli by TE neurons following V4 and TEO lesions in macaques. *Cereb. Cortex*, *15*(2), 141–151.

Cameron, E. L., Tai, J. C., & Carrasco, M. (2002). Covert attention affects the psychometric function of contrast sensitivity. *Vis. Res.*, *42*(8), 949–967.

Carrasco, M., Loula, F., & Ho, Y. X. (2006). How attention enhances spatial resolution: Evidence from selective adaptation to spatial frequency. *Percept. Psychophys.*, *68*(6), 1004–1012.

Carrasco, M., Giordano, A. M., & McElree, B. (2004). Temporal performance fields: Visual and attentional factors. *Vis. Res.*, *44*(12), 1351–1365.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nat. Neurosci.*, *2*(7), 671–676.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol.*, *80*(6), 2918–2940.

Clark, V. P., Parasuraman, R., Keil, K., Kulansky, R., Fannon, S., Maisog, J. M., et al. (1997). Selective attention to face identity and color studied with fMRI. *Hum. Brain Mapp.*, *5*, 293–297.

Cook, E., & Maunsell, J. (2002). Attentional modulation of behavioral performance and neuronal responses in middle temporal and ventral intraparietal areas of macaque monkey. *J. Neurosci.*, *22*(5), 1994–2004.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discrimination of shape, color, and speed: Functional anatomy by positron emission tomography. *J. Neurosci.*, *11*(8), 2383–2402.

Crick, F. H. C. (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proc. Natl. Acad. Sci. USA*, *81*(14), 4586–4590.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, *18*, 193–222.

Driver, J., Baylis, G. C., & Rafal, R. D. (1992). Preserved figure-ground segregation and symmetry perception in visual neglect. *Nature*, *360*, 73–75.

Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.*, *113*(4), 501–517.

Freiwald, W. A., & Kanwisher, N. G. (2004). Visual selective attention: Insights from brain imaging and neurophysiology. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (3rd ed., pp. 575–588). Cambridge, MA: MIT Press.

Haenny, P. E., Maunsell, J. H., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Exp. Brain Res.*, *69*(2), 245–259.

Han, S., Jiang, Y., Mao, L., Humphreys, G. W., & Gu, H. (2005). Attentional modulation of perceptual grouping in human visual cortex: Functional MRI studies. *Hum. Brain Mapp.*, *25*(4), 424–432.

Hayden, B. Y., & Gallant, J. L. (2005). Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4. *Neuron*, *47*, 637–643.

Heinze, H. J., Luck, S. J., Munte, T. F., Gos, A., Mangun, G. R., & Hillyard, S. A. (1994). Attention to adjacent and separate positions in space: An electrophysiological analysis. *Percept. Psychophys.*, *56*(1), 42–52.

Hopf, J. M., Boehler, C. N., Luck, S., Tsotsos, J. K., Heinze, H. J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proc. Natl. Acad. Sci. USA*, *103*(4), 1053–1058.

Kanizsa, G. (1976). Subjective contours. *Sci. Am.*, *234*, 48–52.

Kapadia, M., Ito, M., Gilbert, C., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *15*, 843–856.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, *282*(5386), 108–111.

Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *J. Neurophysiol.*, *86*(3), 1398–1411.

Kastner, S., De Weerd, P., & Ungerleider, L. G. (2000). Texture segregation in the human visual cortex: A functional MRI study. *J. Neurophysiol.*, *83*, 2453–2457.

Kastner, S., Nothdurft, H., & Pigarev, I. (1999). Neuronal responses to motion and orientation contrast in cat striate cortex. *Visual Neurosci.*, *16*, 587–600.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761.

Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.*, *67*(4), 961–980.

Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.*, *15*, 1605–1615.

Lamy, D., Segal, H., & Ruderman, L. (2006). Grouping does not require attention. *Percept. Psychophys.*, *68*(1), 17–31.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Percept. Psychophys.*, *56*(2), 183–197.

Lee, J., Williford, T., & Maunsell, J. H. (2007). Spatial attention and the latency of neuronal responses in macaque area V4. *J. Neurosci.*, *27*(36), 9632–9637.

Li, Z. (1999). Contextual influences in V1 as a basis for pop-out and asymmetry in visual search. *Proc. Natl. Acad. Sci. USA*, *96*(18), 10530–10535.

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vis. Res.*, *38*(9), 1183–1198.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.*, *77*, 24–42.

Mack, A., Tang, B., Tuma, R., Kahn, S., & Rock, I. (1992). Perceptual organization and attention. *Cogn. Psych.*, *24*(4), 475–501.

Marcus, D. S., & Van Essen, D. (2002). Scene segmentation and attention in primate cortical areas V1 and V2. *J. Neurophysiol.*, *88*(5), 2648–2658.

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.*, *2*(4), 364–369.

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.*, *14*(9), 744–751.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci.*, *29*(6), 317–322.

McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.*, *19*(1), 431–441.

McAdams, C. J., & Maunsell, J. H. (2000). Attention to both space and feature modulates neuronal responses in macaque area V4. *J. Neurophysiol.*, *83*(3), 1751–1755.

McMains, S. A., Fehd, H. M., Emmanouli, T. A., & Kastner, S. (2007). Mechanisms of feature- and space-based attention: Response modulation and baseline increases. *J. Neurophysiol.*, *98*(4), 2110–2121.

McMains, S. A., & Kastner, S. (2007). Illusory contour formation modulates competitive interactions in human extrastriate cortex. *J. Vis.*, *7*(9), abstract 112, 112a.

Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2000). Intermodal selective attention in monkeys. I. Distribution and timing of effects across visual areas. *Cereb. Cortex*, *10*(4), 343–358.

Miller, E. K., Gochin, P. M., & Gross, C. G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Res.*, *616*(1–2), 25–29.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron*, *55*(1), 131–141.

Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *J. Exp. Psychol. Hum. Percept. Perform.*, *23*(2), 339–352.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, *421*(6921), 370–373.

Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.*, *14*(4), 2178–2189.

Muller, N. G., & Kleinschmidt, A. (2003). Dynamic interaction of object- and space-based attention in retinotopic visual areas. *J. Neurosci.*, *23*(30), 9812–9816.

Nothdurft, H.-C., Gallant, J. L., & Van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: Correlates of "popout" under anesthesia. *Visual Neurosci.*, *16*, 15–34.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nat. Neurosci.*, *5*(11), 1203–1209.

O'Craven, K., Downing, P., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, *18*(4), 591–598.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychon. Bull. Rev.*, *1*(1), 29–55.

Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. *Cogn. Psychol.*, *24*(3), 436–447.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Pinsk, M. A., Doniger, G. M., & Kastner, S. (2004). Push-pull mechanism of selective attention in human extrastriate cortex. *J. Neurophysiol.*, *92*(1), 622–629.

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.*, *32*, 3–25.

Qiu, F. T., Sugihara, T., & von der Heydt, R. (2007). Figure-ground mechanisms provide structure for selective attention. *Nat. Neurosci.*, *10*(10), 1492–1499.

Recanzone, G. H., & Wurtz, R. H. (2000). Effects of attention on MT and MST neuronal activity during pursuit initiation. *J. Neurophysiol.*, *83*(2), 777–790.

Recanzone, G. H., Wurtz, R. H., & Schwarz, U. (1997). Responses of MT and MST neurons to one and two moving objects in the receptive field. *J. Neurophysiol.*, *78*(6), 2904–2915.

Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science*, *278*(5343), 1616–1619.

Ress, D., Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nat. Neurosci.*, *3*(9), 940–945.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.*, *19*(5), 1736–1753.

Reynolds, J. H., & Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron*, *37*(5), 853–863.

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.*, *29*, 203–227.

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–381.

Rubin, E. (1958). Figure ground. In D. C. Beardslee (Ed.), *Readings in perception* (pp. 194–203). Princeton, NJ: Van Nostrand.

Russell, C., & Driver, J. (2005). New indirect measures of "inattentive" visual grouping in a change-detection task. *Percept. Psychophys.*, *67*(4), 606–623.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.*, *5*(7), 631–632.

Schwartz, S., Vuilleumier, P., Hutton, C., Maravita, A., Dolan, R. J., & Driver, J. (2005). Attentional load and sensory competition in human vision: Modulation of fMRI responses by load at fixation during task-irrelevant stimulation in the peripheral visual field. *Cereb. Cortex*, *15*(6), 770–786.

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, *55*(2), 301–312.

Serences, J. T., Yantis, S., Culberson, A., & Awh, E. (2004). Preparatory activity in visual cortex indexes distractor suppression during covert spatial orienting. *J. Neurophysiol.*, *92*(6), 3538–3545.

Sherman, S. M., & Guillery, R. W. (2001). *Exploring the thalamus*. San Diego: Academic Press.

Shiu, L. P., & Pashler, H. (1995). Spatial attention and vernier acuity. *Vis. Res.*, *35*(3), 337–343.

Shulman, G. L., Ollinger, J. M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Petersen, S. E., et al. (1999). Areas involved in encoding and applying directional expectation to moving objects. *J. Neurosci.*, *19*(21), 9480–9496.

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *J. Neurosci.*, *11*(9), 2768–2785.

Sohn, W., Chong, S. C., Papathomas, T. V., & Vidnyanszky, Z. (2005). Cross-feature spread of global attentional modulation in human area MT+. *NeuroReport*, *16*(12), 1389–1393.

Sylvester, C. M., Shulman, G. L., Jack, A. I., & Corbetta, M. (2007). Asymmetry of anticipatory activity in visual cortex predicts the locus of attention and perception. *J. Neurosci.*, *27*(52), 14424–14433.

Tootell, R., Hadjikhani, N., Hall, E., Marrett, S., Vanduffel, W., Vaughan, J., et al. (1998). The retinotopy of visual spatial attention. *Neuron*, *21*, 1409–1422.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psych.*, *12*, 97–136.

Treue, S., & Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579.

Wertheimer, M. (1923). Laws of organization in perceptual forms. In *A source book of Gestalt psychology*. London: W. Ellis (1938).

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.*, *15*(3), 419–433.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*(6706), 72–75.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.*, *20*(17), 6594–6611.

14 The Frontoparietal Attention Network

MAURIZIO CORBETTA, CHAD M. SYLVESTER, AND GORDON L. SHULMAN

ABSTRACT This chapter is concerned with the attentional mechanisms that ensure that behavior is directed toward important stimuli in the environment. We review the evidence for a coherent neural network in dorsal parietal and frontal cortex that sends top-down signals, reflecting both the location and features of task-relevant objects, that bias processing in sensory regions such as occipital cortex. Top-down signals for location aid selection of an object by changing neural activity throughout an occipital retinotopic map, not just at the attended location, resulting in a relative increase in activity at that location in the map. The overall behavioral goals that determine which objects are selected are not set within the frontoparietal network, but reflect the interaction of networks involved in reward, memory, and executive control. These networks may provide inputs to dorsal frontoparietal regions that are transformed into biasing signals.

"Attention" broadly refers to a set of mechanisms that allow people to selectively perceive and respond to events that are relevant to their behavioral goals. Because of the importance of this function for many aspects of human behavior, discussions of attention crop up in treatments of such diverse topics as language, memory, emotion, perception, and motor control. Here we concentrate on the selection of objects in the environment for action. Chapter 13 in this book (Kastner, McMains, & Beck) discussed how attending to an object biases sensory evoked activity in sensory cortex and how these biases are thought to result in selective processing of the object (Desimone & Duncan, 1995).

Because complex goal-directed behaviors reflect the interaction of many different brain systems, it is not possible to speak of a single attentional control system. Different brain networks may be recruited when formulating behavioral goals, assessing those goals with respect to current knowledge of the environment, retrieving relevant information from memory, and integrating all of these influences into specific biasing signals that can be sent down to sensory and motor systems (hence the term "top-down" biases). In this review, we suggest that a set of dorsal frontoparietal regions constitute a *dorsal frontoparietal attention network* that performs this final integration step. While this network operates irrespective of the criteria used to select stimuli (e.g., location, features) or responses (e.g., effector), subregions may show specializations for particular attributes, as we will discuss. We will also briefly review the coordination of this network with other networks involved in assessing value, generating and maintaining goals, accessing information in working memory, and retrieving information from long-term memory (figure 14.1).

MAURIZIO CORBETTA, CHAD M. SYLVESTER, and GORDON L. SHULMAN Departments of Neurology, Radiology, and Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri

The dorsal frontoparietal attention network

DEFINITION: FUNCTIONAL CONNECTIVITY AND ANTICIPATORY SIGNALS While the involvement of different brain regions in different functions is not controversial, it may not seem justified to segregate sets of regions into coherent brain networks, such as a dorsal frontoparietal network. However, an important development over the last decade has been the refinement of physiological techniques for identifying related sets of brain regions. One such technique, called functional connectivity magnetic resonance imaging (fcMRI), measures the temporal correlation of the blood-oxygenation-level dependent (BOLD) signal across multiple regions (Biswal, Yetkin, Haughton, & Hyde, 1995). Related regions show strong low-frequency (< 0.1 Hz) correlations over time, even when the subject is lying at rest with no task or stimulation (resting-state fcMRI). The origin of these correlations is still controversial, but they are thought to reflect both anatomical and functional factors. Several studies in the last five years have identified a number of resting-state networks that correspond to regions that are coactivated when subjects perform a task (Damoiseaux et al., 2006; Fox, Corbetta, Snyder, Vincent, & Raichle, 2006; Fox et al., 2005; Fransson, 2005; Greicius, Krasnow, Reiss, & Menon, 2003; Hampson, Peterson, Skudlarski, Gatenby, & Gore, 2002). Relevant to this discussion is the strong correlation between the frontal eye field (FEF) at the intersection of superior frontal sulcus and precentral sulcus, and regions within the intraparietal sulcus (IPS). IPS and FEF represent the core regions of the dorsal attention network (figure 14.2). These regions show spontaneous correlation also with visual areas like

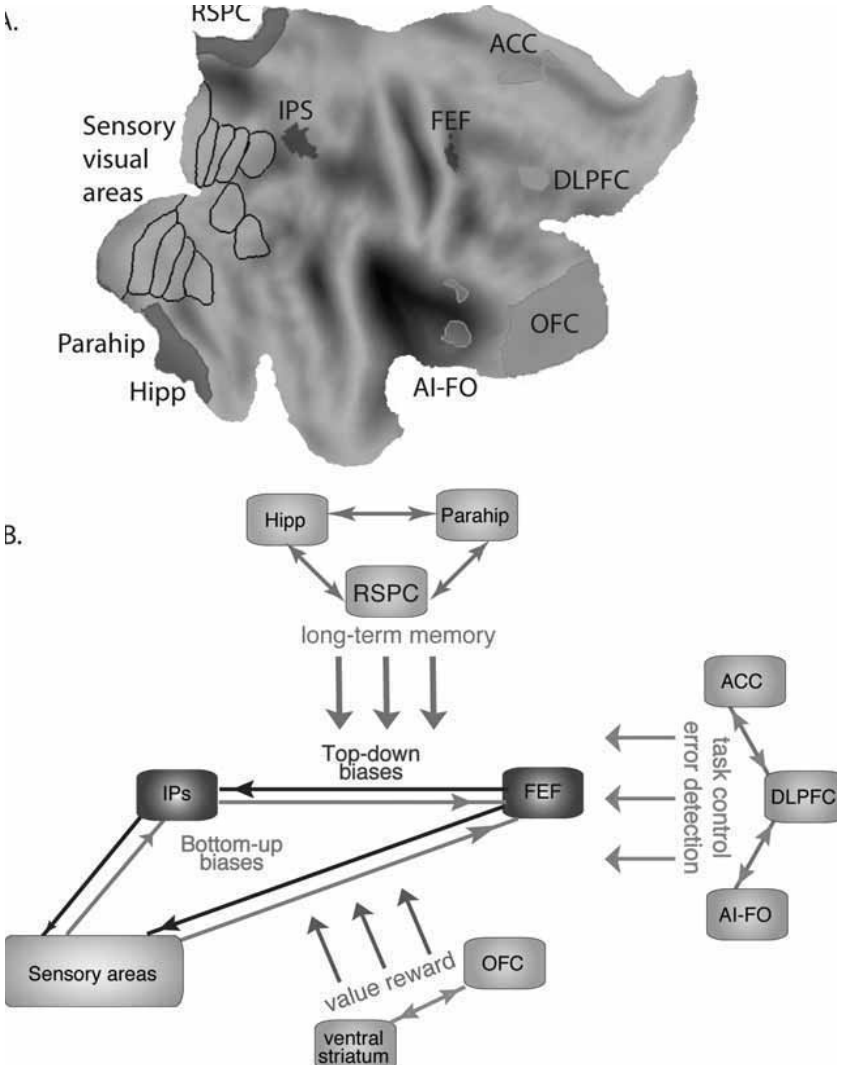


FIGURE 14.1 (*A*) Control and data-processing networks. Flat map of right hemisphere on which regions and different networks involved in control are superimposed. Dark blue: dorsal frontoparietal attention network. IPS: intraparietal sulcus; FEF: frontal eye field. Orange: sensory visual areas. Purple: Long-term memory retrieval network. RSPC: retrosplenial cortex; Parahip: parahippocampus; Hipp: hippocampus. Azure: executive control network. ACC: anterior cingulate; DLPFC: dorsolateral prefrontal cortex; AI-FO: anterior insula–frontal operculum. Green: reward value network. OFC: orbitofrontal cortex; ventral striatum (not shown). (*B*) Wire diagram. Dorsal attention network feeds top-down and receives bottom-up biases to/from sensory cortices for stimulus and response selection. Other networks bias sensory processing via interaction with dorsal attention network. (See color plate 14.).

MT+ and V7. Other networks involved in regulation of attention are shown in figure 14.2 and will not be further considered in this chapter. A right-hemisphere-dominant ventral frontoparietal attention network, with core regions in right temporoparietal junction and ventral frontal cortex, is involved in stimulus-driven reorienting and resetting task-relevant networks; its physiological properties have been recently reviewed (Corbetta, Patel, & Shulman, 2008). A bilateral "default" network is consistently deactivated during goal-directed behavior (Mazoyer et al., 2001; Raichle et al., 2001; Shulman et al., 1997) and may be important in filtering information from internal task-irrelevant processes.

Perhaps the most direct method for observing sensory biases in isolation is to provide subjects with a cue telling them to attend to a specific location in space or a visual feature and to measure the resulting physiological signals prior to the onset of a target stimulus (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Hopfinger, Buonocore, & Mangun, 2000; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; N. Muller, Bartelt, Donner, Villringer, & Brandt, 2003; Serences, Yantis, Culberson, & Awh, 2004; Sylvester, Shulman, Jack, & Corbetta, 2007). Cuing experiments in humans have routinely observed preparatory or endogenous activations—that is, activations not

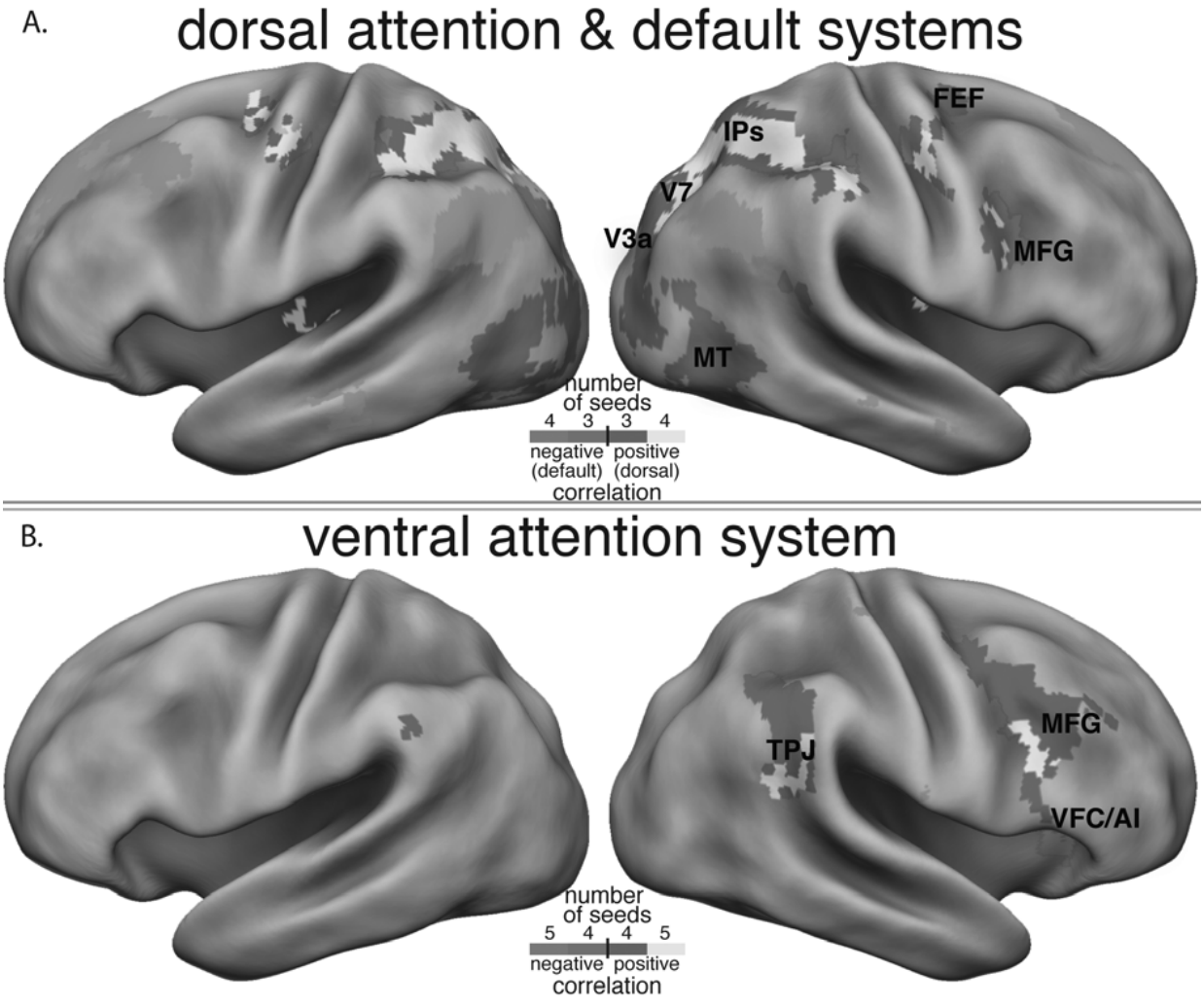
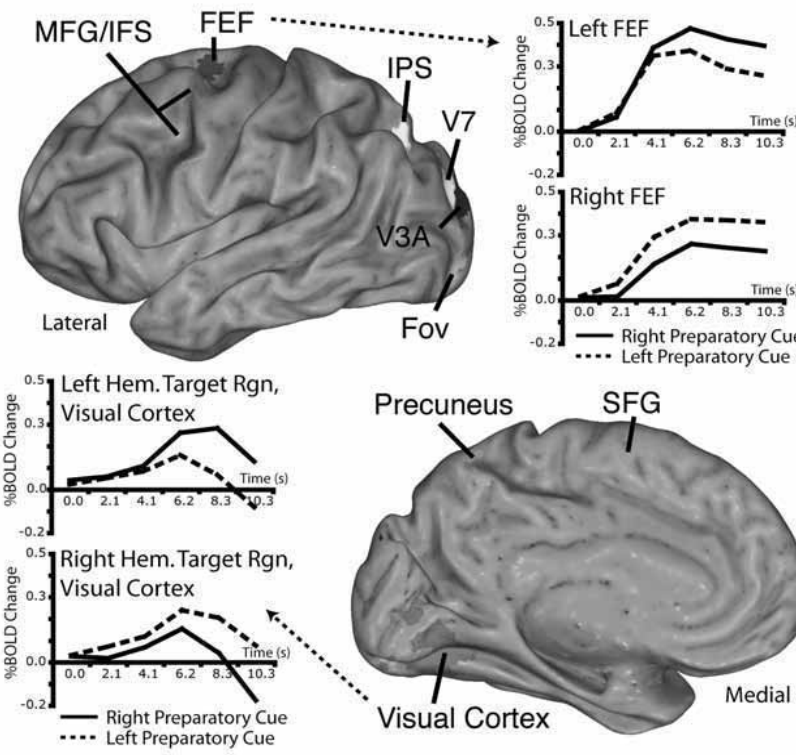


FIGURE 14.2 Functional connectivity by fMRI (fcMRI) defines separate dorsal and ventral networks. (*A*) Dorsal attention and default networks. The map indicates regions that showed significant positive correlations with three (red) or four (yellow) of the seed regions in the dorsal attention network (IPS, FEF, V7, MT+). The dorsal network is largely reproduced in the resting state FC maps. Regions that show significant negative correlations with three (green) or four (blue) of the seed regions are also shown and roughly reproduce the default network, possibly indicating a push-pull relationship between the two networks. (*B*) Ventral attention network. Five ventral regions (R TPJ, R VFC, R MFG, R PrCe) were used as seeds for an FC analysis. Regions showing consistent positive correlations largely reproduce the ventral network, but negative correlations in default regions are not observed. The posterior MFG near the inferior frontal sulcus appears to be connected to both networks. (He et al., 2007.) (See color plate 15.)

driven by a sensory stimulus—in dorsal parietal regions of IPS, extending medially into superior parietal lobule, and in dorsal precentral sulcus at the intersection with the superior frontal sulcus (FEF; figure 14.3*A*). As noted earlier, these regions also show strong resting-state fcMRI. Dorsal frontoparietal activations are observed whether the cue stimulus is visual (Corbetta et al., 2000; Hopfinger et al.; Kastner et al.) or auditory (Sylvester et al.) and are sustained as attention is maintained over extended durations (Corbetta, Kincade, & Shulman, 2002). Subtle but consistent topographic differences have been reported between regions that encode a cue and maintain attention (Woldorff et al., 2004). Preparatory activations in parietal and other regions predict performance on subsequent targets (Pessoa & Padmala, 2005; Sapir, d'Avossa, McAvoy, Shulman, & Corbetta, 2005). Finally, purely endogenous activations in dorsal frontoparietal regions are spatially selective, with greater activity following a cue in contralateral FEF and IPS (Sylvester et al.), as expected if these regions control the spatial selection of information.

In some studies of spatial cuing, endogenous cue-related responses in dorsal frontoparietal regions are accompanied by spatially selective endogenous activation of retinotopic occipital cortex (Kastner et al., 1999; Sylvester et al., 2007),

A.


MFG/IFS
FEF
IPS
V7
V3A
Lateral
Fov
Left FEF
Right FEF
%BOLD Change
Time (s)
0.0 2.1 4.1 6.2 8.3 10.3
Right Preparatory Cue
Left Preparatory Cue
Left Hem. Target Rgn, Visual Cortex
Right Hem. Target Rgn, Visual Cortex
Precuneus
SFG
Medial
Visual Cortex

B.

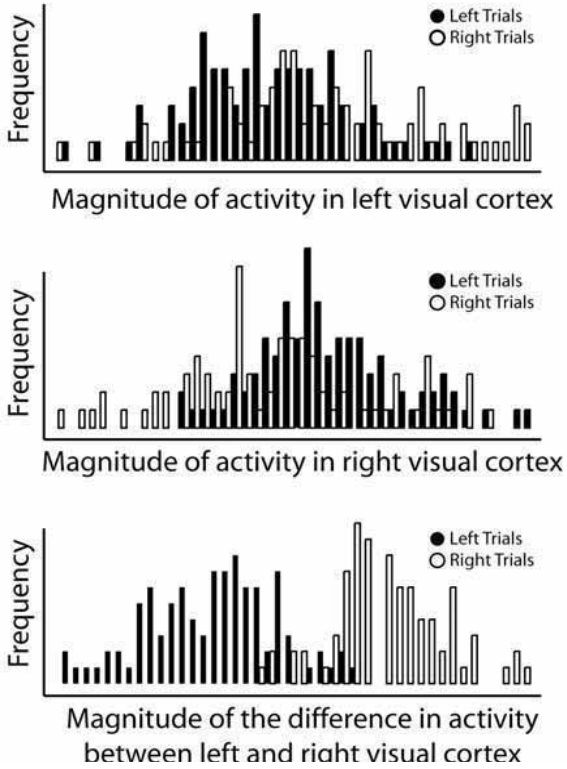

Frequency
Left Trials
Right Trials
Magnitude of activity in left visual cortex
Magnitude of activity in right visual cortex
Magnitude of the difference in activity between left and right visual cortex

FIGURE 14.3 (*A*) Frontoparietal areas and visual cortex modulated by anticipatory signal for spatial attention. Areas with spatially selective preparatory signals following an auditory cue directing attention to a left or right location. MFG/IFS: middle frontal gyrus/inferior frontal sulcus; FEF: frontal eye field; IPS: intraparietal sulcus; Fov: foveal region of V1–V3; SFG: superior frontal gyrus. BOLD signal time series following spatial auditory cues show anticipatory signals that are stronger for cue directing attention to contralateral visual field locations. (See color plate 16.) (*B*) Relative activity predicts locus of attention in visual cortex. Top: single-trial response magnitude of anticipatory activity in left visual cortex for directing spatial attention to left or right visual location. Note slightly higher activity for contralateral attention (right trials), but also strong overlap in the two populations of trials. Therefore readout of activity from only attended visual cortex does not provide a good prediction of the locus of spatial attention. Middle: same for right visual cortex. Bottom: the difference in activity between left and right visual cortex provides a strong trial-to-trial prediction of the locus of spatial attention by subtracting out common noise in the two maps. (From Sylvester et al., 2007.)

but not in other studies (Corbetta, Tansy et al., 2005; figure 14.3*A*). Although the reasons for this variation are not well understood, it may reflect the degree to which selection of an object or stimulus is limited by perceptual factors, which may be associated with endogenous modulation of visual cortex, as opposed to factors related to memory or stimulus-response translation.

Therefore dorsal frontal and parietal regions in IPS and FEF that are coactivated by cues to attend to a visual object form a distinct network in resting-state fcMRI studies (Fox et al., 2006, 2005) and can be considered a separate functional network. Interestingly, consistent with the activation results, this network is not correlated under resting conditions with regions in the occipital lobe, except for human MT+. The interaction of dorsal frontoparietal cortex with occipital cortex is highly task contingent. In a later section, we discuss a possible mechanism for flexibly changing the effects of signals in one area (e.g., FEF) on those in another (e.g., V4). These human imaging studies of preparatory and resting-state activity in dorsal frontoparietal regions in humans are complemented by monkey single-unit studies showing anticipatory signals for spatial attention in FEF (Kodaka, Mikami, & Kubota, 1997) and LIP (Bisley & Goldberg, 2003), as well as resting-state functional connectivity between FEF and LIP in anesthetized monkeys (Vincent et al., 2007).

CAUSALITY OF TOP-DOWN BIASES FROM DORSAL ATTENTION NETWORK ONTO VISUAL CORTEX Physiological studies are generally correlational, demonstrating a relationship between a spatial or temporal pattern of neural activity and some task or behavioral parameters. Although cues to attend can produce endogenous signals in both dorsal frontoparietal and sensory cortex, and these signals may be predictive of behavioral performance, these results do not imply a causal influence of control regions on data-processing regions. Several recent studies, however, have provided evidence for this proposition. Moore and colleagues showed in monkey that stimulation of R FEF modulated sensory-evoked activity in V4 neurons whose receptive fields matched the movement field of the stimulated FEF site (Moore & Armstrong, 2003). Stimulation also changed the psychophysical threshold for detection of a low-contrast stimulus presented within the movement field of the stimulated site in FEF (Moore & Fallah, 2001). In humans, Ruff and colleagues showed that TMS of human FEF produced BOLD activation in peripheral V1–V4, independently of whether a stimulus was present (Ruff et al., 2006; figure 14.4*A*). Correspondingly, stimulation enhanced the perceived contrast of peripheral stimuli. These studies show that activity in FEF can produce the physiological changes in occipital cortex and the changes in behavioral performance that are expected for a brain region involved in top-down control of spatial attention.

However, these results were not obtained under physiological conditions, and more importantly they do not demonstrate an asymmetry in the interaction between occipital and dorsal frontoparietal regions. A recent study using Granger causality analysis meets both these objections (Bressler, Tang, Sylvester, Shulman, & Corbetta, 2008). The authors compared the degree to which endogenous preparatory signals in occipital cortex temporally predicted signals in dorsal frontoparietal regions over and above the prediction based on the dorsal frontoparietal regions themselves, and vice versa. They found that signals in dorsal frontoparietal regions strongly predicted occipital signals (top-down direction), and this prediction was significantly greater than that from occipital signals to dorsal frontoparietal signals (bottom-up direction). In fact, the latter predictability did not exceed a baseline level of predictability between any two voxels in the brain. These results support the hypothesis that under physiological conditions, following a cue to attend to a location, dorsal frontoparietal regions modulate occipital regions (figure 14.1).

TOPOGRAPHIC ORGANIZATION OF MAPS IN DORSAL ATTENTION NETWORK To understand how control is implemented, a helpful hint is the functional organization of an area or system, that is, the parameters that are coded in the pattern of neural activity. Recent studies have begun to detail the organization of the spatially selective neurons in frontal and parietal cortex that may be the source of top-down biases to visual cortex. Human parietal and frontal cortex appears to contain topographic maps of contralateral

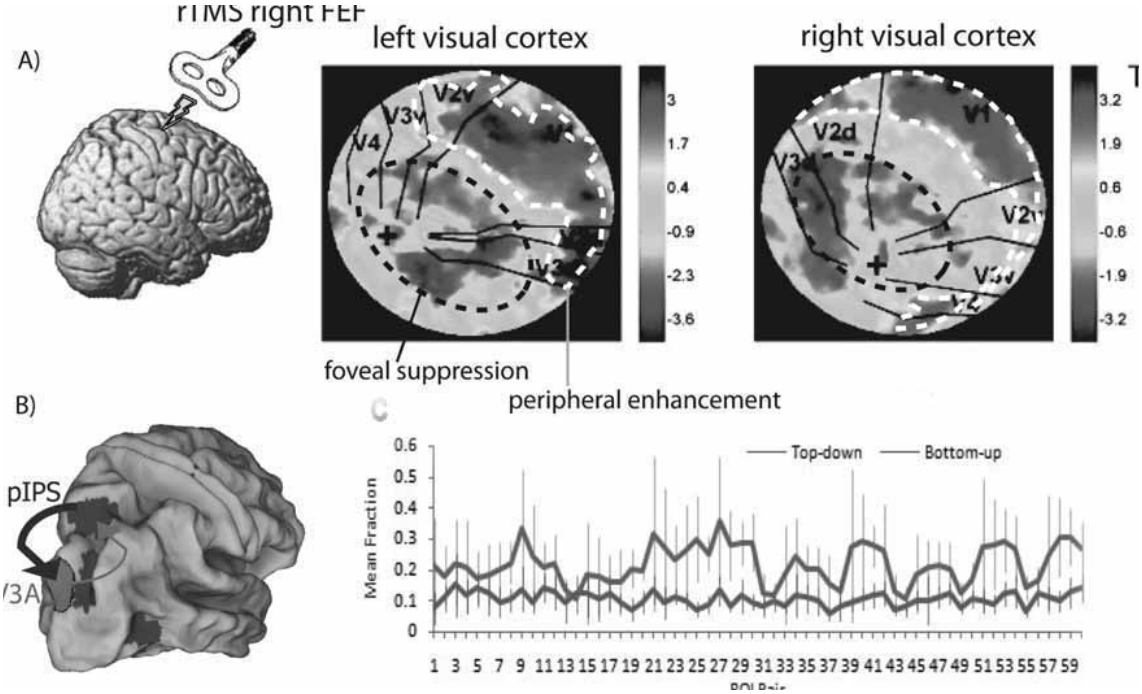


FIGURE 14.4 (*A*) Top-down activation of visual cortex from frontal eye field. Repetitive transcranial magnetic stimulation (rTMS) of right FEF causes bilateral activation in peripheral retinotopic visual cortex (V1–V4) (dotted white line) and corresponding deactivation in foveal representation (dotted black line). (Courtesy of Ruff & Driver; Ruff et al., 2006.) (*B*) Granger causality of BOLD signal time series during anticipatory spatial attention. *Left*: "Control" area pIPS top-down modulates visual area V3A. The strength of top-down control trial-by-trial correlates with higher accuracy on a difficult visual discrimination task. *Right*: During anticipatory spatial attention the strength of top-down influences is stronger from control to visual areas than bottom-up influences from visual areas to control areas. (From Bressler, Tang, Sylvester, Shulman, & Corbetta, 2008.) (See color plate 17.)

visual space. The initial report of a single topographic map of the contralateral hemifield in human parietal cortex by Sereno, Pitzalis, and Martinez (2001) has been followed by several studies that have found multiple maps (Hagler, Riecke, & Sereno, 2007; Schluppeck, Glimcher, & Heeger, 2005; Silver, Ress, & Heeger, 2005), including a report of five contiguous maps along IPS (Swisher, Halko, Merabet, McMains, & Somers, 2007). Unlike early retinotopic occipital cortex, some of these regions may also show substantial nontopographic activations to ipsilateral stimuli (Jack et al., 2007). In the animal single-unit literature, the evidence for topographic maps in parietal areas such as LIP is inconsistent (Ben Hamed, Duhamel, Bremmer, & Graf, 2001; Platt & Glimcher, 1998), although a recent monkey fMRI study reported clear evidence for a hemifield map in both hemispheres (Gaurav Patel, Larry Snyder, and Maurizio Corbetta, personal communication).

Models of spatial selection propose that objects in a scene are represented in a topographically organized "salience" map and that the most activated object or location in the map is selected as part of a "winner-take-all" process (Koch & Ullman, 1985; Wolfe, 1994). The salience of an object in the map is partly determined by its sensory properties (e.g., high contrast) and by whether its features or location are relevant to the current task and, as such, have been biased by preparatory activity. Selection is therefore based on the magnitude of an object's activation relative to the activation of all other objects in the scene. A single-unit study in area LIP (Bisley & Goldberg, 2003), a likely homologue of human IPS regions, supports the idea that selection is based on the activity in one spatially selective set of neurons relative to that of another. The duration for which a monkey attended to the location of a flashed distracter object, presented in the opposite hemifield from an attended target location, was predicted from the duration for which LIP activity from the neurons responding to the distracter location was greater than the activity from neurons responding to the target location. In other words, the locus of attention in frontoparietal control regions may be coded by a difference signal between the level of anticipatory activity at the attended location versus unattended locations in different parts of the map.

MECHANISMS: CODING THE LOCUS OF ATTENTION BASED ON RELATIVE ACTIVITY WITHIN A MAP The importance of relative rather than absolute activity extends to top-down biases in visual cortex, where attending to a location changes activity not only at the attended location

of the retinotopic visual map but also throughout the map. While preparatory increases are observed at the attended location in an occipital retinotopic map (Hopfinger et al., 2000; Kastner et al., 1999; N. Muller et al., 2003; Serences, Yantis, et al., 2004; Sylvester et al., 2007), preparatory decreases are observed at unattended locations in the map (Silver, Ress, & Heeger, 2007; Sylvester, Jack, Corbetta, & Shulman, 2008).

Mapwide changes in occipital areas support the hypothesis that the operative spatial signal determining the locus of attention and salience is relative, such as a difference signal, but more direct evidence has recently been reported. Most studies of attentional modulations compare the activations from an object when it is attended versus unattended. The assumption is that the two situations are analogous to when attended and unattended objects are simultaneously present. However, this assumption is only correct when the activations from the two objects are uncorrelated over time. In fact, a recent fMRI study has shown that preparatory signals between homotopic locations of occipital retinotopic maps, as well as between left and right IPS and FEF, are highly correlated over trials (Sylvester et al., 2007). Correlated activity is significant but less strong at nonhomotopic locations (e.g., between fovea and periphery) or across separate areas (e.g., between FEF and V3A). Correlated neural activity across separate parts of V1 following stimulus presentation has also been reported (Chen, Geisler, & Seidemann, 2006), suggesting that the BOLD correlations reflect neural rather than or in addition to hemodynamic factors. While the neural causes of the BOLD correlations are not known, they might reflect nonspatial signals that carry information about the upcoming stimulus features or task or overall changes in arousal. As a result of the correlated signal, the absolute BOLD signal at the attended location in occipital and dorsal frontoparietal cortex is only a moderate trial-to-trial predictor of the direction of attention. Predictability in both occipital and dorsal frontoparietal areas is greatly improved by subtracting out the common "noise," that is, taking the difference between activity at the attended location in the map and the homotopic location in the opposite hemisphere (Sestieri et al., 2008; Sylvester et al., 2007; figure 14.3*B*). The biological relevance of the difference signal is demonstrated by the fact that performance for subsequent targets is better predicted in V3A by the magnitude of the difference signal than by the magnitude of the absolute signal at the attended location.

The importance of relative rather than absolute signal levels qualifies the association of sustained BOLD activity with the maintenance of attention (Silver, Ress, & Heeger, 2007). In one fMRI study, while the absolute signal at the cued location in occipital retinotopic maps decreased over the course of the cue period, the difference between the signal at that location and at the homotopic location, as well as the predictability of the locus of attention, increased over the same period, reflecting the larger signal decreases at the homotopic uncued location (Sylvester et al., 2007). Therefore, the maintenance of attention may be reflected in the sustained magnitude of a relative signal, not an absolute signal.

Finally, the mapwide distribution of preparatory activity in retinotopic cortex reflects not simply the location of the attended object, but also the computational demands of the task. Serences, Yantis, and colleagues (2004) demonstrated that when subjects expected a target stimulus to be surrounded by closely spaced distracter objects, preparatory signals increased, even when task difficulty remained constant. However, the relationship between the additional preparatory signal at the attended location and the suppression of distractor information at nontarget locations was unclear. A recent study (Sylvester et al., 2008) has shown that preparatory activity at nontarget locations depends on whether noise at those locations can adversely affect performance. Sylvester and colleagues compared mapwide changes in activity when subjects performed a coarse orientation discrimination on a low-contrast, near-threshold Gabor patch and an equally difficult but fine orientation discrimination on a high-contrast, suprathreshold Gabor patch. The location of the Gabor was cued, and expected contrast was blocked, allowing subjects to optimally adjust the distribution of attention to both the attended location and the nature of the discrimination. In the low-contrast condition, performance was partly limited by noise at nontarget locations that created spurious false alarms, but in the high-contrast condition, performance was only limited by noise at the target location. Even though the spatial distribution of the task stimulus was identical in the two conditions, preparatory signal decreases at nontarget locations in retinotopic occipital maps were greater when subjects expected a low-contrast than high-contrast Gabor, reflecting the need to suppress noise at these locations and "mark" the target location by creating a steep target-nontarget location gradient. No effects of expected contrast were observed at the cued location.

How are these task-dependent changes in the mapwide distribution of preparatory signals controlled? Sylvester and colleagues (2008) reported that regions in FEF and IFS (inferior frontal sulcus), but not IPS, showed additive effects of expected target location and contrast, with greater activations when contralateral locations were cued but also when low-contrast stimuli were expected at either contralateral or ipsilateral locations. The additive contrast and cue location signals in FEF and IFS were combined to produce the interacting, mapwide changes observed in retinotopic occipital cortex, although the manner in which this process occurred was unclear. Interestingly, stimulation of FEF by TMS, in the absence of a visual stimulus, decreases activity in portions of early visual cortex corresponding to the central

visual field (Ruff et al., 2006), which also showed the largest decreases in Sylvester and colleagues (2008).

In conclusion, these studies emphasize that selection of an object at a location in the visual field involves a mapwide modulation of topographic sensory representation, and that spatial sensory bias at the attended location arises from an interaction of biasing signals at both attended and unattended locations. Within dorsal frontoparietal areas, spatial biases coexist with nonspatial (feature) biases that seem to independently modulate activity in visual areas prior to stimulus presentation. The presence of correlated noise, especially among homologous locations of a map (e.g., left and right upper visual field) or regions of a network (e.g., left and right FEF) may reflect spontaneous oscillatory activity that is strongest between regions in visual cortex that are connected through the corpus callosum. In the next section we link these interhemispheric interactions in spatial selection to the functional relationship that exists between spatial attention and the control of eye movements.

Mechanisms: Relationship to Eye Movements and Feature-Based Selection While spatial selection can be controlled independently of eye movements (Klein, 1980), it is striking that the same set of brain regions, including IPS and FEF (and superior colliculus subcortically), have been implicated in both processes (Beauchamp, Petit, Ellmore, Ingeholm, & Haxby, 2001; Bisley & Goldberg, 2003; Bruce & Goldberg, 1985; Bushnell, Goldberg, & Robinson, 1981; Corbetta et al., 1998; Corbetta, Miezin, Shulman, & Petersen, 1993; Luna et al., 1998; Nobre et al., 1997; Perry & Zeki, 2000; Petit, Clark, Ingeholm, & Haxby, 1997; Snyder, Batista, & Andersen, 1997; Sweeney et al., 1996). Moreover, many studies have suggested that both processes are linked (Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995; Kustov & Robinson, 1996; Sheliga, Riggio, & Rizzolatti, 1994; Shepherd, Findlay, & Hockey, 1986; but see Klein, 1980), leading to "premotor" (Rizzolatti, Riggio, Dascola, & Umiltá, 1987) accounts of spatial selection. The basic idea is that the allocation of spatial attention in the visual field, as well as the related effects on behavioral performance and evoked responses in sensory cortex (see chapter 13), depend on the preparation of an eye movement plan to foveate the target. Since most experiments of spatial attention are conducted while subjects maintain fixation, these oculomotor plans are "covert" and do not lead to an actual eye movement. Conversely, the preparation of an eye movement plan leads to an anticipatory shift of spatial attention. This idea has been more recently supported by electrical microstimulation studies in oculomotor regions (FEF and superior colliculus) showing that stimulation with electricity below the level necessary to produce an overt eye movement still produces changes in visual field sensitivity consistent with a shift of spatial attention at the potential saccade location (Moore & Fallah, 2001; J. Muller, Philiastides, & Newsome, 2005). The coincidence of neural systems for spatial attention and eye movements is not complete, however; at the level of single cells in FEF, visual and visuomovement cells respond during covert target selection while purely motor cells are either silent or even inhibited (Thompson, Biscoe, & Sato, 2005). Therefore, a reasonable conclusion is that spatial attention and eye movement mechanisms are functionally (even causally) related but that dissociations can occur both psychologically and neurally (Awh, Armstrong, & Moore, 2006).

The link between spatial attention and eye movements provides a potential explanation for the prominent role that interhemispheric interactions play in the allocation of attention in both visual cortex and frontoparietal regions. As earlier discussed, in both control and sensory areas the locus of spatial attention is best described by a relative balance of activity between attended and unattended parts of a topographic map, especially between homologous regions in different hemispheres (e.g., left and right upper visual field, or left and right FEF). If directing attention is akin to planning an eye movement, then it would be sensible to have mechanisms that prevent the simultaneous activation of oculomotor neurons in opposite directions. These would ensure behavioral coherence as subjects explore the environment. An inhibitory mechanism that prevents reexploration of previously attended locations or objects is well described and is also linked to eye movements (inhibition of return). Here we propose that a similar mechanism may be operative during the anticipatory deployment of attention to a location and the subsequent readout of sensory information. By using a subtraction strategy between homologous regions for coding the locus of attention, the brain prevents the generation of eye movements and selection of stimuli in opposite locations. It will be interesting to evaluate whether similar subtractive strategies are operational even when competing locations are in the same visual field or along the vertical axis.

The important link between attention and eye movements does not completely explain the broad and flexible range of selection mechanisms that are implemented in the dorsal frontoparietal attention network. Selection of visual information is based on many properties besides spatial location. People can efficiently search for an object defined by a very large number of different features (e.g., red, motion in a particular direction) in the presence of complete spatial uncertainty (Wolfe, 1994). Consistent with this observation, feature-specific modulations of sensory activity have been reported throughout the visual field, not just at the currently attended location, both in single-unit and imaging studies (Martinez-Trujillo & Treue, 2004; Saenz, Buracas, & Boynton, 2002; Treue & Trujillo, 1999). Objects containing the primed features, however, attract shifts of attention (Folk, Remington, & Johnston, 1992; Serences et al., 2005),

suggesting that feature-based selection may operate in part by biasing sensory activity to an object, which evokes shifts of spatial attention or eye movements to the object (Shih & Sperling, 1996). This interaction suggests important links between or within brain regions involved in feature-based selection and overt or covert spatial selection.

Accordingly, feature-selective selection signals have been found in parietal and FEF cortex. Fast-latency single-unit responses to a particular color, for example, have been reported in FEF following extended experience with targets defined by that color in multiobject displays (Bichot, Schall, & Thompson, 1996). Imaging studies have reported dorsal frontoparietal activity during cuing of features in addition to spatial location (Shulman et al., 1999). Explicit comparisons of the activations to cues for location and color indicate common activity in many dorsal frontoparietal regions, but also activations in subregions that are greater for location than color (Giesbrecht, Woldorff, Song, & Mangun, 2003; Slagter et al., 2007). Similarly, studies comparing motion and color cues have reported dorsal frontoparietal activations that are greater for motion than color (Mangun & Fannon, 2007; Shulman, D'Avossa, Tansy, & Corbetta, 2002), and Mangun and colleagues have suggested that the location-selective regions are similar to the motion-selective regions (Mangun, Fannon, Geng, & Saron, 2009). These studies indicate that preparing to select a visual attribute activates dorsal frontoparietal regions that generalize across visual dimensions as well as subregions that show specificity for some dimensions. Finally, during sustained attention to a visual stimulus, feature-specific attentional modulations—for example, modulations that are specific to a particular direction of motion—have been reported in parietal cortex and FEF, in addition to retinotopic occipital cortex, using multivoxel classification techniques (Serences & Boynton, 2007).

While feature-based and location-based selection can be coordinated, as in the case of visual search discussed previously, feature-based selection need not drive an oculomotor mechanism. Selection by hierarchical scale is one example in which eye movements do not aid changes in selection, yet preparatory activations are observed in IPS, with larger preparatory activations during selection of local scales in left than right IPS (Weissman & Woldorff, 2005). Therefore, while eye movements, feature-based selection, and object-based selection sometimes operate in a coordinated fashion, premotor theories that explain selection of task-relevant visual stimuli solely in terms of the signals that drive or are observed within oculomotor mechanisms remain incomplete. Dorsal frontoparietal regions include both oculomotor-based and non-oculomotor-based selection mechanisms.

Finally, some transient neural signals appear to generalize across virtually any change in selection criteria. Transient "switch" signals are observed most strongly in medial parietal areas such as precuneus, rather than IPS, and in frontal areas such as FEF, whether attention is switched between stimuli in different locations (Yantis et al., 2002), between superimposed objects (Serences, Schwarzbach, Courtney, Golay, & Yantis, 2004), between stimuli in different modalities (Shomstein & Yantis, 2004), or between superimposed random-dot arrays with different visual features (Liu, Slotnick, Serences, & Yantis, 2003). Switch signals have been hypothesized to enable networks to "settle" into a state appropriate to the newly attended object (Serences & Yantis, 2006).

Establishing goals in frontoparietal network

In the laboratory, instructions to direct selective attention are provided by symbolic or nonsymbolic cues (arrows, symbols, sensory cues, etc). In real life, however, selective attention is controlled by a complex combination of signals (goals, desires, memories), which sit in the background of awareness while guiding behavior and are generated in distributed brain systems that interact with the frontoparietal network.

Several classes of internal signals can drive selective attention. If we take the simple case of "reaching for a cup in the cupboard to drink some water," orienting to the cupboard and selecting the appropriate action to reach for the cup engages over time the frontoparietal dorsal attention network. But this preparatory activity reflects the interaction of the dorsal network with other neural systems. For instance, this behavior is motivated by an error signal in the hypothalamus indicating a mismatch in the concentration of blood (or osmolarity), which we subjectively perceive as thirst. The overall organization of the behavior ("reach for the cup in the cupboard") may be built from learned patterns that can be loaded in working memory. Signals from long-term memory indicate the kitchen's layout and the position of the cups in the cupboard. Therefore a network controlling spatial attention and target selection should interact over time with other neural systems monitoring the internal milieu and expected reward, working memory, and long-term memory. The actual sequence of activity is currently unknown because of the lack of suitable methods for tracking the temporal evolution of neural activity in distributed neural networks. However, the available evidence does show that the spatial attention system is jointly activated with other cognitive systems (working memory/executive control; reward; long-term memory) under conditions in which these systems provide inputs for the spatial selection of objects and responses. This idea is presented in figure 14.1, and some of the evidence is presented in the following sections.

Reward/Value Signals and the Limbic System

Neurons in areas involved in the control of spatial attention and eye movements carry information about the amount of

reward a monkey receives for making an instructed eye movement. Platt and Glimcher found that the activity of neurons in the lateral intraparietal (LIP) area coded not only the position of the visual stimulus or intended eye movement but also the probability that a particular eye movement response would yield a fruit juice reward and the expected amount of the reward (Platt & Glimcher, 1999). Moreover, when animals were left to make their own decision on where to look, neuronal activity varied with the probability and size of an upcoming fruit juice reward. Reward-related modulations have also been observed in several areas connected to LIP like superior colliculus (Dorris & Munoz, 1998) and prefrontal cortex (Gold & Shadlen, 2003). To the extent that LIP and related areas contain salience maps indicating locations that are likely targets for eye movements or shifts of attention, the salience map includes information concerning the expected value of the eye movement.

While activity in the dorsal attention network combines information about expected value and the physical location or attributes of the object that can guide a shift of attention or eye movement, other neural systems code only reward information. Reward-related signals involve a distributed network of brain stem and cortical regions including sensory-motor and limbic regions. Two recent reports showed that neurons in orbitofrontal cortex coded the value of two possible choices independently of their physical attributes or locations, unlike LIP (Padoa-Schioppa & Assad, 2006, 2008). Other regions that have been proposed within the limbic system as the putative source of motivational/reward signals are the nucleus accumbens and related frontal connections, amygdala, and anterior and posterior cingulate cortex.

The posterior cingulate, in particular, is strongly connected bidirectionally with both limbic areas and sensory parietal areas, and may funnel limbic influences onto regions involved in visuospatial attention, as originally proposed by Marsel Mesulam (1981). Several human brain-imaging experiments have manipulated spatial attention and reward. In one study subjects directed attention based on central cues, and in different blocks won or lost money depending on whether their response to visual targets was faster or slower than following a neutral cue (Small et al., 2005). Posterior cingulate, anterior cingulate, orbitofrontal, and parahippocampal cortex were modulated by the interaction of attention-based facilitation of performance and reward. In a second study, subjects were asked to direct spatial attention based on central cues to identify target stimuli that were either food or nonfood items (Mohanty, Gitelman, Small, & Mesulam, 2008). In one session subjects were hungry, whereas in another session they were satiated. Activity in orbitofrontal cortex, posterior cingulate, and regions of the dorsal attention system (IPS) was more correlated with performance benefits due to spatial attention during hunger than satiety. Single-unit studies have also demonstrated reward-related modulations in posterior cingulate neurons of macaques. Activity was correlated positively and negatively, in different neurons, with the size of a reward and was also affected by the omission of an expected reward (McCoy, Crowley, Haghighian, Dean, & Platt, 2003).

Task Sets, Prefrontal-Cingular Circuits, and Working Memory Task sets and working-memory signals also guide stimulus selection. Informally, a task set specifies the types of stimuli that should be mapped onto particular responses, although more formal definitions have been developed (Logan & Gordon, 2001). A task set can be open ended, such as "get a cup in a cupboard for drinking," or more constrained, like "get a red cup among green and blue cups."

Experimental evidence indicates that prefrontal cortex, through its widespread connection with other brain areas, plays a critical role in generating, maintaining, and applying a task set. Patients with prefrontal lesions have problems carrying out complex behaviors, especially when they require multiple steps, and are prone to distraction by irrelevant environmental stimuli (Fuster, 1985; Stuss & Benson, 1986). Neurons in prefrontal cortex respond to cues that instruct a task, maintain this information throughout a temporal delay in the absence of sensory stimulation, and can prospectively code for the relevant response (Miller & Cohen, 2001). Prefrontal cortex is especially important when cues guiding behavior must be flexibly changed over time (Rossi, Bichot, Desimone, & Ungerleider, 2007).

Task sets that are represented in prefrontal cortex may determine the particular stimulus-response linkages or sensory biases established by the dorsal attention network. A distinction between orienting and executive networks was originally proposed in a seminal review by Posner and Petersen (1990). Recent studies have provided more evidence for a functional separation of task control and stimulus selection systems. In a large fMRI meta-analysis, Dosenbach and colleagues isolated two different putative control networks: an anterior cingulofrontal operculum (ACC/FO) network and a frontoparietal network (Dosenbach et al., 2007, 2006). The ACC/FO network was defined based on the presence across studies of three putative control signals: (1) a sustained signal maintained throughout a block of trials (typically 30–50 seconds), possibly representing a "sustained" set signal for the task at hand; (2) a transient signal at the beginning and end of a block of trials, possibly representing the loading and unloading of task parameters; and (3) a transient signal generated by an error, possibly underlying error correction and on-line adjustments. The co-localization of these three signals across a large number of different tasks to the same cingular-opercular regions supports the hypothesis that these regions form a functional network involved in some aspect of high-

level control. Moreover, the same regions formed a network in a resting-state fcMRI study that was distinct from a frontoparietal network. Dosenbach and colleagues suggested that the latter network, which only partially overlapped the dorsal frontoparietal network involved in stimulus selection, was involved in moment-to-moment task adjustments. While the functional analysis of the cingulo-opercular network is just beginning and its role in behavioral performance is not well understood, these results support a functional subdivision between a dorsal frontoparietal attention network and a high-level ACC/FO network that maintains an abstract specification of a task.

During task performance, task information is thought to be maintained in "working memory," a short-term storage in which information can be easily manipulated and accessed. The neural structures underlying visual working memory have been investigated under conditions in which this task information is highly constrained and specific (e.g., the position of each cup in the cupboard), with parietal regions showing load-dependent activity and a correlation of this activity with memory performance (Marois, Chun, & Gore, 2004; Todd, Fougnie, & Marois, 2005; Vogel, McCollough, & Machizawa, 2005). Early studies showed a substantial overlap of frontoparietal regions involved in maintaining spatial attention at a location and maintaining spatial working memory (Cabeza & Nyberg, 1997; Corbetta, Kincade, & Shulman, 2002), with behavioral results clearly showing that one is important for the other (Awh & Jonides, 2001).

More recent studies by Nobre and colleagues indicate overlapping mechanisms for orienting to objects in spatial working memory and orienting to objects in the environment (Lepsien, Griffin, Devlin, & Nobre, 2005; Nobre et al., 2004). In a "pre-cue" environment condition, a cue directed attention to a location before the onset of an array. In a working-memory "retro-cue" condition, subjects loaded the same array into working memory, and then saw a cue indicating a location in the array. In both cases subjects decided whether a test object was present in the stimulus array. Both pre- and retro-cues facilitated the identification of objects that were respectively presented in the visual field or in working memory, and correspondingly, the standard frontoparietal network was also recruited in both conditions. However, in addition, "retro-cues" activated to a greater degree medial and lateral prefrontal cortex, while "pre-cues" activated visual cortex prior to stimulus presentation. Prefrontal activation precedes posterior activation in the memory condition (Lepsien et al.). These findings suggest, as in the case of reward and motivational signals discussed earlier, that prefrontal regions putatively involved in working memory coactivate with the dorsal frontoparietal network when spatial attention is directed to a memory representation.

Long-Term Memory and Hippocampus Long-term memory signals also guide spatial orienting under ecological conditions. Once a goal is established, prior knowledge about the surrounding environment guides the execution of complex behavior in a nearly effortless manner. In reaching for a cup, we direct attention to the location of the cabinet based on prior knowledge about the layout of the kitchen as well as the location of a particular cup in the cabinet.

Behavioral studies have shown that implicit memory derived from previous exposure to a particular stimulus configuration can facilitate performance during visual search of a target among distractors. These results indicate influences of long-term memory on orienting as we move and act in the environment (Chun & Jiang, 2003). There is also evidence that semantic knowledge, such as the association of a word with an object, can facilitate the detection of that object (Moores, Laiti, & Chelazzi, 2003).

A recent study showed that attention can be directed to specific objects in a visual scene based on memory and that the frontoparietal attention network is recruited when attention is guided by memory similarly to the way in which it is guided by explicit visual cues (Summerfield, Lepsien, Gitelman, Mesulam, & Nobre, 2006). Summerfield and colleagues asked subjects to learn the location of a target object in several visual scenes. The next day they were asked to detect the same target object in previously learned or novel scenes. Performance was facilitated when subjects knew where to look based on prior experience or a visual cue. Interestingly, the time course of this memory-based facilitation was quite rapid (~100 ms), consistent with an automatic deployment of attention. The frontoparietal attention network was recruited similarly in the memory and visual cuing conditions, whereas in the memory condition additional regions related to memory retrieval were engaged (parahippocampal, retrosplenial, hippocampus). Interestingly, the degree of hippocampus activation across subjects correlated with the strength of the spatial facilitation in the memory condition, consistent with a large literature in rats associating the hippocampus and related structures to spatial memory (O'Keefe, Burgess, Donnett, Jeffery, & Maguire, 1998; Wilson & Tonegawa, 1997).

Physiological mechanisms for top-down influences and selection of visual objects

The preceding evidence indicates that the dorsal frontoparietal attention network for stimulus selection is recruited under different task conditions, and that correspondingly different neural systems are coactivated with that network. Although coactivation does not imply a functional interaction, it seems likely that distributed networks underlying

reward, executive control, working memory, and long-term memory all influence the biasing signals sent from the dorsal frontoparietal network to sensory cortex.

The issue of communication between distributed neural networks is a fundamental problem in neuroscience that is particularly important in discussions of attention. While there are several potential mechanisms for linking separate distant neuronal populations, one recent idea is that information is best transmitted from one group of neurons to another; that is, neural activity in one group of neurons most strongly affects activity in a second group of neurons when their neural activity is synchronized or coherent (Fries, 2005; see also chapter 20 by Womelsdorf and Fries). Therefore, synchronization could be a powerful mechanism for controlling whether signals from sensory areas are sent to higher-order areas, that is, whether an object is selected. There is now a substantial and growing body of evidence that rhythmic synchronization is not epiphenomenal. A number of studies have investigated *local* neuronal synchronization and demonstrated its relation to, for example, selective attention (Bichot, Rossi, & Desimone, 2005; Engel, Fries, & Singer, 2001; Taylor, Mandon, Freiwald, & Kreiter, 2005; Womelsdorf, Fries, Mitra, & Desimone, 2006) and working-memory maintenance (Howard et al., 2003; Pesaran, Pezaris, Sahani, Mitra, & Andersen, 2002).

Local neuronal synchronization in parietal, frontal, and occipital cortex during visual attention to a stimulus has been reported at both low (8–15 Hz) and high (40–80 Hz) frequency. A number of EEG/MEG studies in healthy subjects have shown anticipatory decreases in alpha/beta power over parieto-occipital cortex before stimulus presentation that predicts subsequent visual discrimination (Babiloni, Vecchio, Miriello, Romani, & Rossini, 2006; Sauseng et al., 2005; Thut, Nietzel, Brandt, & Pascual-Leone, 2006). Interestingly, alpha/beta power modulations have an asymmetrical topography, with greater decrements in the hemisphere contralateral to the side of attention (in the case of spatial attention), and the degree of asymmetry is more predictive of the locus of attention and performance than the absolute power modulation in either hemisphere in isolation. These asymmetrical anticipatory changes in EEG power are highly reminiscent of the asymmetrical BOLD anticipatory changes in frontoparietal and visual regions during spatial attention tasks (see the first section, "The dorsal frontoparietal attention network"). As noted earlier, the best predictor of the locus of attention and subsequent perception to targets is the difference between anticipatory activity in frontoparietal and visual regions contralateral and ipsilateral to the side of attention. Preliminary evidence indicates a functional link between BOLD and EEG signal fluctuations in the alpha/beta band at rest in the dorsal frontoparietal attention network (Laufs et al., 2003; Mantini, Perrucci, Del Gratta, Romani, & Corbetta, 2007).

Conclusions

There is now consensus in cognitive neuroscience that dorsal frontoparietal areas are the sources of top-down biases onto data-processing areas in sensory cortices. The most important anticipatory signal that these areas provide is information about the upcoming target location, but there is growing evidence for other "object"- or "identity"-related signals being coded in these regions, perhaps in conjunction with other parts of prefrontal cortex. Perhaps the most important problem for the future is to find ways to visualize and thus understand how these intermediate control regions interact with other systems involved in coding value/reward, long-term memory, and goals. Solving this problem will necessarily require higher temporal resolution methods coupled with methods like fMRI that provide enough spatial resolution to monitor neural activity at the level of single areas. Another important goal will be to find the computations or the set of transformations to which these regions contribute, given the generality of their recruitment in many forms of selection. Will higher specificity emerge when population-level activity from these areas is monitored at higher temporal/spatial resolution? The difficulty of these problems, however, should not preclude us from using functional anatomical models like the one presented here and elsewhere (Corbetta et al., 2008; Corbetta & Shulman, 2002) to begin understanding human brain diseases. A network approach to the study of disorders of attention has in fact already yielded critical information (Corbetta, Kincade, Lewis, Snyder, & Sapir, 2005; Posner, Walker, Friedrich, & Rafal, 1984) that will be used in the near future to improve human health.

REFERENCES

Awh, E., Armstrong, K. M., & Moore, T. (2006). Visual and oculomotor selection: Links, causes and implications for spatial attention. *Trends Cogn. Sci.*, *10*(3), 124–130.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends Cogn. Sci.*, *5*(3), 119–126.

Babiloni, C., Vecchio, F., Miriello, M., Romani, G. L., & Rossini, P. M. (2006). Visuo-spatial consciousness and parieto-occipital areas: A high-resolution EEG study. *Cereb. Cortex*, *16*(1), 37–46.

Beauchamp, M. S., Petit, L., Ellmore, T. M., Ingeholm, J., & Haxby, J. V. (2001). A parametric fMRI study of overt and covert shifts of visuospatial attention. *Neuroimage*, *14*(2), 310–321.

Ben Hamed, S., Duhamel, J. R., Bremmer, F., & Graf, W. (2001). Representation of the visual field in the lateral intraparietal area of macaque monkeys: A quantitative receptive field analysis. *Exp. Brain Res.*, *140*(2), 127–144.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*(5721), 529–534.

Bichot, N. P., Schall, J. D., & Thompson, K. G. (1996). Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature*, *381*(6584), 697–699.

Bisley, J. W., & Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, *299*(5603), 81–86.

Biswal, B., Yetkin, F., Haughton, V., & Hyde, J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.*, *34*, 537–541.

Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., & Corbetta, M. (2008). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J. Neurosci*, *28*(40), 10056–10061.

Bruce, C. J., & Goldberg, M. E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *J. Neurophysiol.*, *53*, 603–635.

Bushnell, M. C., Goldberg, M. E., & Robinson, D. L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *J. Neurophysiol.*, *46*(4), 755–772.

Cabeza, R., & Nyberg, L. (1997). Imaging cognition: An empirical review of PET studies with normal subjects. *J. Cogn. Neurosci.*, *9*, 1–26.

Chen, Y., Geisler, W. S., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nat. Neurosci.*, *9*(11), 1412–1420.

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *J. Exp. Psychol. Learn. Mem. Cogn.*, *29*(2), 224–234.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*, 761–773.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.*, *3*, 292–297.

Corbetta, M., Kincade, J. M., & Shulman, G. L. (2002). Neural systems for visual orienting and their relationship with working memory. *J. Cogn. Neurosci.*, *14*(3), 508–523.

Corbetta, M., Kincade, M. J., Lewis, C., Snyder, A. Z., & Sapir, A. (2005). Neural basis and recovery of spatial attention deficits in spatial neglect. *Nat. Neurosci.*, *8*(11), 1603–1610.

Corbetta, M., Miezin, F. M., Shulman, G. L., & Petersen, S. E. (1993). A PET study of visuospatial attention. *J. Neurosci.*, *13*(3), 1202–1226.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*(3), 306–324.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.*, *3*(3), 201–215.

Corbetta, M., Tansy, A. P., Stanley, C. M., Astafiev, S. V., Snyder, A. Z., & Shulman, G. L. (2005). A functional MRI study of preparatory signals for spatial location and objects. *Neuropsychologia*, *43*(14), 2041–2056.

Damoiseaux, J. S., Rombouts, S. A., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., et al. (2006). Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. USA*, *103*(37), 13848–13853.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, *18*, 193–222.

Dorris, M. C., & Munoz, D. P. (1998). Saccadic probability influences motor preparation signals and time to saccadic initiation. *J. Neurosci.*, *18*(17), 7015–7026.

Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proc. Natl. Acad. Sci. USA*, *104*(26), 11073–11078.

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., et al. (2006). A core system for the implementation of task sets. *Neuron*, *50*(5), 799–812.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.*, *2*(10), 704–716.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol. Hum. Percept. Perform.*, *18*(4), 1030–1044.

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl. Acad. Sci. USA*, *103*(26), 10046–10051.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. USA*, *102*(27), 9673–9678.

Fransson, P. (2005). Spontaneous low-frequency BOLD signal fluctuations: An fMRI investigation of the resting-state default mode of brain function hypothesis. *Hum. Brain Mapp.*, *26*(1), 15–29.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn. Sci.*, *9*(10), 474–480.

Fuster, J. M. (1985). The prefrontal cortex and temporal integration. In E. G. Jones & A. Peters (Eds.), *Cerebral Cortex* (Vol. 4, pp. 151–177). New York: Plenum Press.

Giesbrecht, B., Woldorff, M. G., Song, A. W., & Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *NeuroImage*, *19*(3), 496–512.

Gold, J. I., & Shadlen, M. N. (2003). The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *J. Neurosci.*, *23*(2), 632–651.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. USA*, *100*(1), 253–258.

Hagler, D. J., Jr., Riecke, L., & Sereno, M. I. (2007). Parietal and superior frontal visuospatial maps activated by pointing and saccades. *NeuroImage*, *35*(4), 1562–1577.

Hampson, M., Peterson, B. S., Skudlarski, P., Gatenby, J. C., & Gore, J. C. (2002). Detection of functional connectivity using temporal correlations in MR images. *Hum. Brain Mapp.*, *15*(4), 247–262.

He, B. J., Snyder, A. Z., Vincent, J. L., Epstein, A., Shulman, G. L., & Corbetta, M. (2007). Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron*, *53*(6), 905–918.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Percept. Psychophys.*, *57*, 787–795.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat. Neurosci.*, *3*, 284–291.

Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., et al. (2003). Gamma oscillations correlate with working memory load in humans. *Cereb. Cortex*, *13*(12), 1369–1374.

Jack, A. I., Patel, G. H., Astafiev, S. V., Snyder, A. Z., Akbudak, E., Shulman, G. L., et al. (2007). Changing human visual field organization from early visual to extra-occipital cortex. *PLoS One*, *2*, e452.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*, 751–761.

Klein, R. (1980). Does oculomotor readiness mediate cognitive control of visual attention? In R. S. Nickerson (Ed.), *Attention and performance VII* (pp. 259–276). Hillsdale, NJ: Erlbaum.

Koch, C., & Ullman, S. (1985). Shifts in visual attention: Towards the underlying circuitry. *Hum. Neurobiol.*, *4*, 219–227.

Kodaka, Y., Mikami, A., & Kubota, K. (1997). Neuronal activity in the frontal eye field of the monkey is modulated while attention is focused on to a stimulus in the peripheral visual field, irrespective of eye movement. *Neurosci. Res.*, *28*, 291–298.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vis. Res.*, *35*, 1897–1916.

Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *384*, 74–77.

Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., et al. (2003). Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proc. Natl. Acad. Sci. USA*, *100*(19), 11053–11058.

Lepsien, J., Griffin, I. C., Devlin, J. T., & Nobre, A. C. (2005). Directing spatial attention in mental representations: Interactions between attentional orienting and working-memory load. *NeuroImage*, *26*(3), 733–743.

Liu, T., Slotnick, S. D., Serences, J. T., & Yantis, S. (2003). Cortical mechanisms of feature-based attentional control. *Cereb. Cortex*, *13*(12), 1334–1343.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychol. Rev.*, *108*(2), 393–434.

Luna, B., Thulborn, K. R., Strojwas, M. H., McCurtain, B. J., Berman, R. A., Genovese, C. R., et al. (1998). Dorsal cortical regions subserving visually-guided saccades in humans: An fMRI study. *Cereb. Cortex*, *8*, 40–47.

Mangun, G. R., & Fannon, S. P. (2007). Networks for attentional control and selection in spatial vision. In F. Mast & L. Jancke (Eds.), *Spatial processing in navigation, imagery, and perception* (pp. 411–432). Amsterdam: Springer.

Mangun, G. R., Fannon, S. P., Geng, J. J., & Saron, C. D. (2009). Imaging brain attention systems: Control and selection in vision. In M. Filippi (Ed.), *fMRI techniques and protocols*. Totowa, NJ: Humana Press.

Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L., & Corbetta, M. (2007). Electrophysiological signatures of resting state networks in the human brain. *Proc. Natl. Acad. Sci. USA*, *104*(32), 13170–13175.

Marois, R., Chun, M. M., & Gore, J. C. (2004). A common parieto-frontal network is recruited under both low visibility and high perceptual interference conditions. *J. Neurophysiol.*, *92*(5), 2985–2992.

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.*, *14*(9), 744–751.

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houde, O., et al. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res. Bull.*, *54*(3), 287–298.

McCoy, A. N., Crowley, J. C., Haghighian, G., Dean, H. L., & Platt, M. L. (2003). Saccade reward signals in posterior cingulate cortex. *Neuron*, *40*(5), 1031–1040.

Mesulam, M.-M. (1981). A cortical network for directed attention and unilateral neglect. *Ann. Neurol.*, *10*, 309–325.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.*, *24*, 167–202.

Mohanty, A., Gitelman, D. R., Small, D. M., & Mesulam, M. M. (2008). The spatial attention network interacts with limbic and monoaminergic systems to modulate motivation-induced attention shifts. *Cereb. Cortex.*, *18*(11), 2604–2613.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, *421*(6921), 370–373.

Moore, T., & Fallah, M. (2001). Control of eye movements and spatial attention. *Proc. Natl. Acad. Sci. USA*, *98*(3), 1273–1276.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nat. Neurosci.*, *6*(2), 182–189.

Muller, J. R., Philiastides, M. G., & Newsome, W. T. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proc. Natl. Acad. Sci. USA*, *102*(3), 524–529.

Muller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A physiological correlate of the "zoom lens" of visual attention. *J. Neurosci.*, *23*(9), 3561–3565.

Nobre, A. C., Coull, J. T., Maquet, P., Frith, C. D., Vandenberghe, R., & Mesulam, M. M. (2004). Orienting attention to locations in perceptual versus mental representations. *J. Cogn. Neurosci.*, *16*(3), 363–373.

Nobre, A. C., Sebestyen, G. N., Gitelman, D. R., Mesulam, M. M., Frackowiack, R. S. J., & Frith, C. D. (1997). Functional localization of the system for visuospatial attention using positron emission tomography. *Brain*, *120*, 515–533.

O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J., & Maguire, E. A. (1998). Place cells, navigational accuracy, and the human hippocampus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *353*(1373), 1333–1340.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226.

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.*, *11*(1), 95–102.

Perry, R. J., & Zeki, S. (2000). The neurology of saccades and covert shifts in spatial attention: An event-related fMRI study. *Brain*, *123*(Pt. 11), 2273–2288.

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., & Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.*, *5*(8), 805–811.

Pessoa, L., & Padmala, S. (2005). Quantitative prediction of perceptual decisions during near-threshold fear detection. *Proc. Natl. Acad. Sci. USA*, *102*(15), 5612–5617.

Petit, L., Clark, V. P., Ingeholm, J., & Haxby, J. V. (1997). Dissociation of saccade-related and pursuit-related activation in

human frontal eye fields as revealed by fMRI. *J. Neurophysiol.*, *77*, 3386–3390.

Platt, M. L., & Glimcher, P. W. (1998). Response fields of intraparietal neurons quantified with multiple saccadic targets. *Exp. Brain Res.*, *121*(1), 65–75.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*(6741), 233–238.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.*, *13*, 25–42.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *J. Neurosci.*, *4*(7), 1863–1874.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). Inaugural article: A default mode of brain function. *Proc. Natl. Acad. Sci. USA*, *98*(2), 676–682.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1a), 31–40.

Rossi, A. F., Bichot, N. P., Desimone, R., & Ungerleider, L. G. (2007). Top down attentional deficits in macaques with lesions of lateral prefrontal cortex. *J. Neurosci.*, *27*(42), 11306–11314.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J.-D., et al. (2006). Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. *Curr. Biol.*, *16*(15), 1479–1488.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.*, *5*, 631–632.

Sapir, A., d'Avossa, G., McAvoy, M., Shulman, G. L., & Corbetta, M. (2005). Brain signals for spatial attention predict performance in a motion discrimination task. *Proc. Natl. Acad. Sci. USA*, *102*(49), 17810–17815.

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., et al. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur. J. Neurosci.*, *22*(11), 2917–2926.

Schluppeck, D., Glimcher, P., & Heeger, D. J. (2005). Topographic organization for delayed saccades in human posterior parietal cortex. *J. Neurophysiol.*, *94*(2), 1372–1384.

Serences, J. T., & Boynton, G. M. (2007). The representation of behavioral choice for motion in human visual cortex. *J. Neurosci.*, *27*(47), 12893–12899.

Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., & Yantis, S. (2004). Control of object-based attention in human cortex. *Cereb. Cortex*, *14*(12), 1346–1357.

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychol. Sci.*, *16*(2), 114–122.

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends Cogn. Sci.*, *10*(1), 38–45.

Serences, J. T., Yantis, S., Culberson, A., & Awh, E. (2004). Preparatory activity in visual cortex indexes distractor suppression during covert spatial orienting. *J. Neurophysiol.*, *92*(6), 3538–3545.

Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, *294*(5545), 1350–1354.

Sestieri, C., Sylvester, C. M., Jack, A. I., d'Avossa, G., Shulman, G. L., & Corbetta, M. (2008). Independence of anticipatory signals for spatial attention from number of nontarget stimuli in the visual field. *J. Neurophysiol.*, *100*(2), 829–838.

Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Exp. Brain Res.*, *98*, 507–522.

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *Q. J. Exp. Psychol.*, *38*, 475–491.

Shih, S., & Sperling, G. (1996). Is there feature-based attentional selection in visual search? *J. Exp. Psychol. Hum. Percept. Perform.*, *22*(3), 758–779.

Shomstein, S., & Yantis, S. (2004). Control of attention shifts between vision and audition in human cortex. *J. Neurosci.*, *24*(47), 10702–10706.

Shulman, G. L., d'Avossa, G., Tansy, A. P., & Corbetta, M. (2002). Two attentional processes in the parietal lobe. *Cereb. Cortex*, *12*(11), 1124–1131.

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks. II. Decreases in cerebral cortex. *J. Cogn. Neurosci.*, *9*, 648–663.

Shulman, G. L., Ollinger, J. M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Petersen, S. E., et al. (1999). Areas involved in encoding and applying directional expectations to moving objects. *J. Neurosci.*, *19*(21), 9480–9496.

Silver, M. A., Ress, D., & Heeger, D. J. (2005). Topographic maps of visual spatial attention in human parietal cortex. *J. Neurophysiol.*, *94*(2), 1358–1371.

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *J. Neurophysiol.*, *97*(1), 229–237.

Slagter, H. A., Giesbrecht, B., Kok, A., Weissman, D. H., Kenemans, J. L., Woldorff, M. G., et al. (2007). fMRI evidence for both generalized and specialized components of attentional control. *Brain Res.*, *1177*, 90–102.

Small, D. M., Gitelman, D., Simmons, K., Bloise, S. M., Parrish, T., & Mesulam, M. M. (2005). Monetary incentives enhance processing in brain regions mediating top-down control of attention. *Cereb. Cortex*, *15*(12), 1855–1865.

Snyder, L. H., Batista, A. P., & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature*, *386*, 167–170.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes*. New York: Raven Press.

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron*, *49*(6), 905–916.

Sweeney, J. A., Mintum, M. A., Kwee, S., Wiseman, M. B., Brown, D. L., Rosenberg, D. R., et al. (1996). Positron emission tomography study of voluntary saccadic eye movement and spatial working memory. *J. Neurophysiol.*, *75*, 454–468.

Swisher, J. D., Halko, M. A., Merabet, L. B., McMains, S. A., & Somers, D. C. (2007). Visual topography of human intraparietal sulcus. *J. Neurosci.*, *27*(20), 5326–5337.

Sylvester, C. M., Jack, A. I., Corbetta, M., & Shulman, G. L. (2008). Anticipatory suppression of nonattended locations in visual cortex marks target location and predicts perception. *J. Neurosci.*, *28*(26), 6549–6556.

Sylvester, C. M., Shulman, G. L., Jack, A. I., & Corbetta, M. (2007). Asymmetry of anticipatory activity in visual cortex predicts the locus of attention and perception. *J. Neurosci, 27*(52), 14424–14433.

Taylor, K., Mandon, S., Freiwald, W. A., & Kreiter, A. K. (2005). Coherent oscillatory activity in monkey area V4 predicts

successful allocation of attention. *Cereb. Cortex*, *15*(9), 1424–1437.

Thompson, K. G., Biscoe, K. L., & Sato, T. R. (2005). Neuronal basis of covert spatial attention in the frontal eye field. *J. Neurosci.*, *25*(41), 9479–9487.

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.*, *26*(37), 9494–9502.

Todd, J. J., Fougnie, D., & Marois, R. (2005). Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentional blindness. *Psychol. Sci.*, *16*(12), 965–972.

Treue, S., & Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*, 575–579.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., et al. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, *447*(7140), 83–86.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503.

Weissman, D. H., & Woldorff, M. G. (2005). Hemispheric asymmetries for different components of global/local attention occur in distinct temporo-parietal loci. *Cereb. Cortex*, *15*(6), 870–876.

Wilson, M. A., & Tonegawa, S. (1997). Synaptic plasticity, place cells and spatial memory: Study with second generation knockouts. *Trends Neurosci.*, *20*(3), 102–106.

Woldorff, M. G., Hazlett, C. J., Fichtenholtz, H. M., Weissman, D. H., Dale, A. M., & Song, A. W. (2004). Functional parcellation of attentional control regions of the brain. *J. Cogn. Neurosci.*, *16*(1), 149–165.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychon. Bull. & Rev.*, *1*, 202–238.

Womelsdorf, T., Fries, P., Mitra, P. P., & Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature*, *439*(7077), 733–736.

Yantis, S., Schwarzbach, J., Serences, J., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., et al. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nat. Neurosci.*, *5*, 995–1002.

15 Spatiotemporal Analysis of Visual Attention

JENS-MAX HOPF, HANS-JOCHEN HEINZE, MIRCEA A. SCHOENFELD, AND STEVEN A. HILLYARD

ABSTRACT The fact that perception of a visual event can be improved by focusing attention upon its spatial location has been documented in numerous experiments going back more than a century. It is also well established that visual attention can be selectively allocated to nonspatial stimulus features or to entire objects as integrated feature ensembles. In the quest to identify the neural mechanisms that underlie this perceptual selectivity in human observers, neuroimaging methods, including noninvasive recordings of event-related potentials (ERPs) and event-related magnetic fields (ERMFs), have provided valuable insights. Contributions from the ERP and ERMF methodologies have been particularly important for revealing the time course and rapid coordination of the underlying selection processes. The ever-expanding body of research in this field has made it abundantly clear that visual attention does not rely upon a unitary neural mechanism, but instead that multiple selection processes cooperate in a flexible manner to guarantee the adaptability of attention to a wide range of circumstances. Here we outline some of the principles underlying the flexible coordination of space-, feature-, and object-based selection processes that have emerged from recent studies, with particular emphasis on the contributions of ERP and ERMF recordings in human observers.

Spatial attention

Event-related potential (ERP) recordings have shown how spatial attention influences sensory processing in the visual cortex when attention is directed to briefly flashed stimuli in one visual field while equivalent stimuli in the opposite field are ignored (for reviews see Hillyard & Anllo-Vento, 1998; Hopfinger, Luck, & Hillyard, 2004; Luck, Woodman, & Vogel, 2000; Mangun, 1995). The general finding has been that attended stimuli elicit enlarged early ERP components in the visual cortex during the interval 80–200 ms after stimulus onset relative to unattended stimuli. These amplitude enhancements typically occur without a significant change in component latencies or scalp topographies, suggesting that spatial attention operates by simply controlling the gain of visual input during early processing stages (Hillyard, Vogel, & Luck, 1998); such a sensory gain control mechanism is consistent with observations from single-unit recordings in monkeys (Lee, Williford, & Maunsell, 2007; Luck, Chelazzi, Hillyard, & Desimone, 1997; Maunsell & Cook, 2002). The earliest consistent ERP modulations produced by spatial attention were found to be amplitude enhancements of the initial positive (P1) and subsequent negative (N1) components of the ERP elicited at latencies of around 80–100 ms and 130–200 ms, respectively (see figure 15.1). While these two components were generally modulated in tandem, there is increasing evidence that they reflect different aspects of attentional selection (reviewed in Hopfinger et al.), with the initial P1 modulation reflecting location selection per se and the subsequent N1-modulation reflecting discriminative processing of the stimulus within the focus of attention (Hopf, Vogel, Woodman, Heinze, & Luck, 2002; Vogel & Luck, 2000). The amplitudes of both the P1 and N1 components have been found to covary with the amount of processing resources that are voluntarily allocated to a spatial location (Handy & Mangun, 1997, 2000; Mangun & Hillyard, 1990). Moreover, reflexive orienting to the location of a nonpredictive cue was found to enhance the P1 amplitude to a subsequent co-localized target, suggesting that at least initially the same modulatory processes of location selection are engaged during voluntary and reflexive orienting (Hopfinger & Mangun, 1998, 2001; Hopfinger & Ries, 2005).

THE PROFILE OF THE SPATIAL FOCUS OF ATTENTION The focus of spatial attention has been likened to a spotlight (Posner, 1980), a zoom lens (Eriksen & Yeh, 1985), or a Gaussian gradient (Downing & Pinker, 1985; Shulman, Wilson, & Sheehy, 1985), which enhances processing of visual stimuli within a circumscribed region of space. While there is general agreement that the size of this attended region may be adjusted voluntarily, it has long been debated whether the spotlight of spatial attention has a unitary "beam" or whether it can be divided flexibly to disparate

JENS-MAX HOPF, HANS-JOCHEN HEINZE, and MIRCEA A. SCHOENFELD Department of Neurology, Otto von Guericke University; Leibniz Institute for Neurobiology, Magdeburg, Germany

STEVEN A. HILLYARD Department of Neuroscience, University of California, San Diego, California

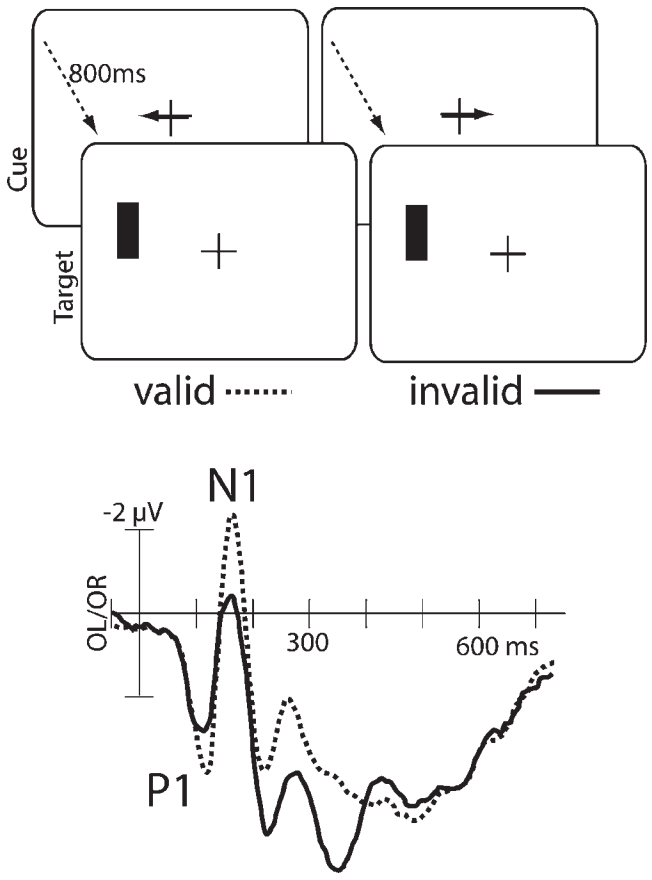


FIGURE 15.1 ERPs recorded in a typical version of Posner's location-cuing paradigm. A small central cue informed subjects of the most probable visual field where the target bar would appear. The subject's task was to perform a demanding length discrimination of the bar. On 75% of the trials the target appeared in the cued VF (valid condition), whereas on 25% of the trials the target appeared in the uncued VF (invalid condition). Below are ERP waveforms elicited by valid (dotted) and invalid targets presented to the same visual field (waveforms collapsed over left and right VF targets). The ERP response to the attended targets shows a prominent amplitude enhancement of the P1 as well as the N1 component without any significant change in latency. (Adapted from Mangun & Hillyard, 1991.)

locations (Awh & Pashler, 2000; Castiello & Umilta, 1992; Hahn & Kramer, 1998; Juola, Bouwhuis, Cooper, & Warner, 1991; Kramer & Hahn, 1995; Posner, 1980). Evidence consistent with the unitary spotlight hypothesis has come from ERP studies in which target arrays were briefly flashed and hence not continuously visible (Eimer, 1999, 2000; Heinze, Luck, et al., 1994). Under conditions where both attended and unattended stimuli were continuously present, however, studies using both electrophysiological (Müller, Malinowski, Gruber, & Hillyard, 2003) and hemodynamic (McMains & Somers, 2004, 2005) measures of attention have found evidence that the spotlight of spatial attention may be divided. Müller, Malinowski, and colleagues (2003) required subjects to attend to two continuously flickering RSVP streams of symbols that were spatially interleaved with two irrelevant streams, each flickering at a different rate. By recording the amplitude of the frequency-tagged steady-state visual evoked potential (SSVEP) to each of the streams, the allocation of attention among the four stimulus positions could be determined. It was found that attending to two spatially separated streams produced enhancement of the SSVEP elicited at those locations but not at an interposed location, indicating that attention had been effectively divided into two separate beams. Similar results were obtained by McMains and colleagues (McMains & Somers, 2004, 2005) in an fMRI study where subjects attended to two out of five rapid serial visual presentation (RSVP) streams of characters, two in each visual quadrant and one at fixation. They observed BOLD enhancements in cortical regions that corresponded retinotopically with the location of the attended streams together with a lack of enhancement in regions corresponding to a spatially interposed distracter stream. Taken together, these physiological studies provide strong evidence that spatial attention may be divided flexibly to noncontiguous zones of the visual field when stimuli are continuously present, as is generally the case in the natural environment.

The concept of a unitary spotlight or zoom lens may be inadequate to describe the spatial distribution of attention even when only one location is attended. The spatial profile of attention is typically envisioned as a simple gradient of enhanced sensory processing that falls off gradually from the center (Downing & Pinker, 1985; Handy, Kingstone, & Mangun, 1996; Henderson & Macquistan, 1993). Some evidence from both ERP and single-unit recordings is consistent with such a profile (Connor, Gallant, Preddie, & Van Essen, 1996; Connor, Preddie, Gallant, & Van Essen, 1997; Eimer, 1997a). However, it has been repeatedly observed that measures of sensory processing may actually be reduced instead of enhanced in the near vicinity of a target item (Bahcall & Kowler, 1999; Cave & Zimmerman, 1997; Cutzu & Tsotsos, 2003; Mounts, 2000; Müller, Mollenhauer, Rosler, & Kleinschmidt, 2005), suggesting that a simple gradient may not represent an accurate picture. Computational models of visual spatial attention such as the Selective Tuning Model (Tsotsos, 2005) have made the explicit prediction of a suppressive zone surrounding the focus of attention. In line with this idea, recent physiological studies have reported observing a more complex center-surround structure. For example, Hopf and colleagues (2006) used MEG recordings to investigate the spatial profile of the focus of attention during visual search for a pop-out target. The ERMF response to a probe stimulus presented at varying distances from the target served as a measure of attentional allocation (figure 15.2*A*). This probe-evoked cortical response was enhanced at the target location but was significantly reduced for probes adjacent to the target (figure 15.2*B*),

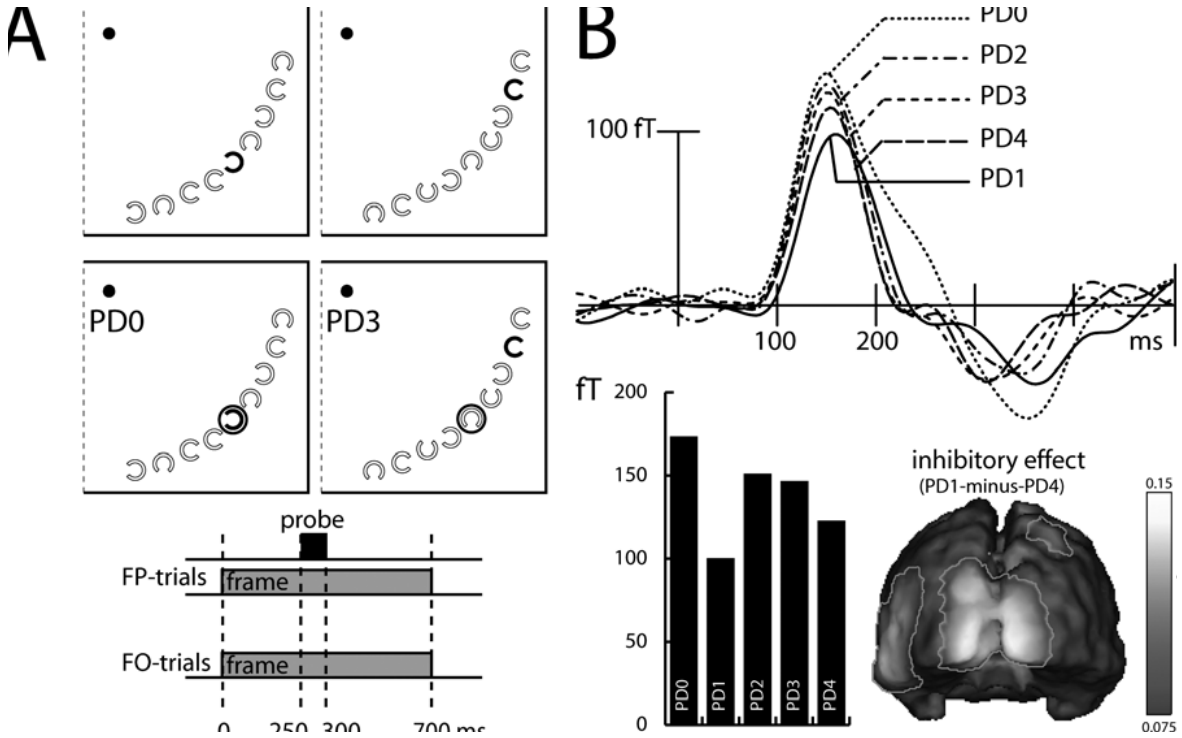


FIGURE 15.2 (*A*) Stimuli used in the study of Hopf and colleagues (2006). Subjects searched for a unique red *C* (shown in black) that randomly appeared at one of nine item locations in the right lower visual quadrant. On 50% of the trials a small white ring (the probe) was presented at the center position 250 ms after search frame onset (frame-probe, or FP, trials). On the remaining trials no probe appeared (frame-only, or FO, trials). On trials with a probe, the target could appear either at the probe's location (probe-distance 0, or PD0, shown on the middle left) or at a location one through four items away from the probe (PD1–PD4) (the situation for PD3 is illustrated on the middle right). The relative timing of search-frame and probe presentation is shown below. (*B*) ERMF response to the probe (FP-minus-FO difference waves) for the five different probe distances (ERMF responses were collapsed across equivalent probe distances toward the horizontal and vertical meridian). A substantial reduction of the probe response was observed when attention was focused next to the probe (PD1) relative to when attention was focused at the probe's location (PD0) or two to four items away (PD2–PD4). This narrow zone of sensory attenuation surrounding the target is further illustrated in the bar graph showing the average size of the probe-related ERMF response between 130 and 150 ms. A source localization analysis of the surround attenuation effect (PD1-minus-PD4 difference) is shown in the lower right, which revealed strongest modulations in early visual cortex areas and smaller effects in lateral and dorsal extrastriate areas. (Adapted from Hopf et al., 2006.)

suggesting that the attended location was surrounded by a narrow zone of relative inhibition. The spatial distribution of attention around a search target thus appears to resemble a Mexican hat profile rather than a simple monotonic gradient. Such a profile may be advantageous in attenuating the most deleterious noise directly adjacent to the target. Studies using ERPs (Slotnick, Hopfinger, Klein, & Sutter, 2002; Slotnick, Schwarzbach, & Yantis, 2003) and fMRI (Müller & Kleinschmidt, 2004) have provided converging evidence and have shown further that surround inhibition may also arise under conditions of sustained focusing.

Locus of Spatial Selection in the Visual System

To specify the anatomical locus of spatial selection in the visual pathways, recordings of the early sensory ERP/ERMF components have been combined with functional brain-imaging methods that provide high spatial resolution. These combined analyses have revealed that the amplitude modulations of the P1 and N1 components produced by spatial attention in the interval 80–200 ms take place in extrastriate cortical areas of both the dorsal and ventral streams (Di Russo, Martinez, & Hillyard, 2003; Heinze, Mangun et al., 1994; Hopf et al., 2002; Mangun, Buonocore, Girelli, & Jha, 1998; Martinez et al., 1999; Noesselt et al., 2002; Woldorff et al., 1997). Notably, in none of these studies were modulatory effects due to attention observed in the initial feedforward sweep of processing in the primary visual cortex, which is reflected in the early ERP/ERMF component known as C1 (50–80 ms) (Aine, Supek, & George, 1995; Clark, Fan, & Hillyard, 1995; Clark & Hillyard, 1996; Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002; Foxe & Simpson, 2002; Olson, Chun, & Allison, 2001). Instead, modulations of V1 activity caused by attention were found to appear after a considerable delay, at latencies of around 150–250 ms, well after the onset of modulatory effects in

extrastriate areas that were reflected in the P1 component (Di Russo et al., 2003; Martinez, Di Russo, Anllo-Vento, Sereno, et al., 2001; Mehta, Ulbert, & Schroeder, 2000a, 2000b; Noesselt et al.). These observations are compatible with the hypothesis that the delayed attention effects in V1 reflect feedback activity originating in higher-level extrastriate areas. This hypothesis is in line with findings from single-unit recordings in monkey primary visual cortex (Motter, 1993; Roelfsema, Lamme, & Spekreijse, 1998; Roelfsema, Tolboom, & Khayat, 2007) and with recent proposals that link the operation of attention with recurrent processing and the sustained integration of information across hierarchical levels in visual cortex (Gilbert & Sigman, 2007; Lamme & Roelfsema, 2000; Shipp, 2007). Along these lines Bullier and colleagues (Bullier, 2001a, 2001b) have suggested that area V1 (and V2) function as an active "blackboard" that serves to integrate and refine visual processing in higher-level areas by means of feedback modulations. Such a refinement may require the persistence of an unbiased and detailed representation of the visual input for at least a short while, until feedback modulatory influences reenter V1/V2.

However, fMRI studies have shown that cue-related anticipatory modulation of neural activity does takes place in striate cortex (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Serences, Yantis, Culberson, & Awh, 2004; Silver, Ress, & Heeger, 2007) as well as in multiple extrastriate areas (Chawla, Rees, & Friston, 1999; Hopfinger, Buonocore, & Mangun, 2000; Müller, Bartelt, Donner, Villringer, & Brandt, 2003) prior to the arrival of the attended stimuli. This anticipatory "biasing" of V1 activity could in principle serve as a mechanism that allows attention to influence the initial feedforward sweep of processing. Consistent with this idea, a recent study by Kelly, Gomez-Ramirez, & Foxe (2008) reported that the visual ERP to a target stimulus presented at a cued location showed an early amplitude increase in the latency range of the C1 component (57–80 ms), which was localized to the vicinity of primary visual cortex. The task design of this experiment differed from those of previous studies in that detection of a low-contrast pattern was required, and this target occurred at a predictable time after an initial cue. If this early ERP modulation does represent an influence of attention over the initial evoked activity in area V1, it would appear that both feedforward and feedback mechanisms in primary visual cortex may be subject to attentional control.

The Role of Spatial Selection in Visual Search In visual search the allocation of spatial attention is strategically different from that seen in cuing tasks, as it is usually contingent on the prior decoding of the target-defining feature(s), which then serve to guide spatial focusing (Wolfe & Horowitz, 2004). Whereas targets having a single unique feature tend to pop out and may be detected and attended rapidly, targets that are defined by a conjunction of two or more features are typically detected with longer latencies that increase as a function of the number of distracters in the display. Some theoretical accounts have proposed that search for a feature conjunction involves the serial focusing of attention on each of the display items (Treisman, 1988), while others have argued for a parallel processing of the multiple object features (Bundesen, 1990; Wolfe, Cave, & Franzel, 1989). This question has been investigated in ERP studies by Woodman and Luck (1999, 2003), which provided strong support for serial search models. Woodman and Luck utilized an ERP component called the N2pc, which is a negative deflection elicited between 180 and 350 ms after search display onset over the posterior cortex contralateral to the target position. The N2pc is well documented to reflect shifts of attention toward the location of the target (Hopf et al., 2000; Hopf, Boelmans, Schoenfeld, Luck, & Heinze, 2004; Luck, Girelli, McDermott, & Ford, 1997; Luck & Hillyard, 1994a, 1994b) and therefore can serve as an "online" measure for tracking rapid spatial shifts of attention. The experimental approach of Woodman and Luck (1999) was to present search displays with two potential target items in opposing visual fields, and then bias search order by introducing bottom-up preferences to focus first on the nontarget item. It was found that the N2pc first appeared over the hemisphere contralateral to the nontarget item and then switched hemispheres as attention was focused upon the target item, suggesting a serial deployment of attention. A further experiment ruled out the possibility that the two potential targets were attended in parallel but with different time courses (Woodman & Luck, 2003), thereby providing unequivocal evidence for the serial allocation of spatial attention during visual search.

Feature-based selection

Beginning with the groundbreaking studies of Corbetta, Miezin, Dobmeyer, Shulman, & Petersen (1990, 1991), it has been well established that paying attention to nonspatial stimulus features results in enhanced neural activity in the cortical areas specialized to process those features (Liu, Slotnick, Serences, & Yantis, 2003; O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997; Saenz, Baracas, & Boynton, 2002; Schoenfeld et al., 2007). For example, a recent study that combined fMRI and neuromagnetic recordings (Schoenfeld et al., 2007) found that a moving stimulus elicited an enhanced neural response in the motion-sensitive area MT when movement was relevant, whereas a color-change stimulus produced greater activity in the color-selective area V4/V8 when color was attended. Paying attention to nonspatial visual features such as color, motion, orientation, or spatial frequency is associated with a class of

ERPs known as the *selection negativities* or *selection positivities*, which are typically enlarged in response to stimuli having the attended feature value (Anllo-Vento, Luck, & Hillyard, 1998; Baas, Kenemans, & Mangun, 2002; Eimer, 1997b; Harter & Aine, 1984; Kenemans, Lijffijt, Camfferman, & Verbaten, 2002; Martinez, Di Russo, Anllo-Vento, & Hillyard, 2001; Wijers, Mulder, Okita, Mulder, & Scheffers, 1989). These feature-selective modulations generally occur at longer latencies (120–300 ms) than the initial effects of spatial attention, but their timing can vary in a flexible manner depending upon feature discriminability and task demands. For example, Schoenfeld and colleagues (2007) found that ERMF and ERP modulations associated with selection *between* feature dimensions (color versus motion) occurred with earlier onset than the modulations typically reported for selection *within* a feature dimension (e.g., one color versus another color).

In an early ERP investigation, Hillyard and Münte (1984) obtained evidence that selection of a relevant stimulus color did not take place outside the spatial focus of attention. In their design, red and blue bars were flashed in random order to both the right and left visual fields, and one color in one field was designated as relevant. It was found that all stimuli in the attended field elicited the enhanced P1/N1 components characteristic of spatial attention, whereas stimuli of the attended color only elicited an enlarged selection negativity in the attended field. This result (confirmed by Anllo-Vento & Hillyard, 1996) suggested that nonspatial feature selection was hierarchically dependent upon the prior selection of location, a notion in line with psychophysical experiments showing that location selection has priority in visual attention (e.g., Cave & Pashler, 1995; Treisman & Gelade, 1980; Tsal & Lamy, 2000).

In contrast to the conclusions of Hillyard and Münte (1984), however, subsequent studies using a variety of methodologies have found that paying attention to a nonspatial stimulus feature does enhance neural responses to that feature even outside the spatial focus of attention. In single-unit recordings from the motion-selective area MT in the monkey, for example, Treue and Martinez-Trujillo (Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999) found that cells tuned to an attended direction of motion showed enhanced firing even when attention was directed to that direction of motion outside the cell's receptive field. The degree of enhancement increased as a function of the similarity between the attended motion direction and the cell's directional preference in a multiplicative way, leading to the proposal that attention operates by increasing the "feature-similarity gain" within the visual cortex (see also Maunsell & Treue, 2006; McAdams & Maunsell, 1999; Motter, 1994). Analogous observations were made using fMRI in human observers (Saenz et al., 2002, 2003). These observers attended to one feature value in a mixed display that was presented to one visual field. In some studies the display was a field of intermingled horizontally and vertically moving dots, and in other studies it was intermingled red and green dots. The general finding was that a larger BOLD signal was elicited in the visual cortex by the attended than the unattended feature value even for stimuli presented to the visual field opposite to where attention was being directed, thereby supporting the concept of "global feature-based attention." A further demonstration that feature-selective attention is not location bound has recently been provided in a study of SSVEPs in humans (Müller et al., 2006). Müller and colleagues presented subjects with superimposed random dot arrays of two colors that flickered at different rates (7.0 and 11.7 Hz) while changing position randomly. The frequency-tagged SSVEP to each color could therefore be measured independently under conditions where location could not be used as a basis for selection. It was found that the color-specific SSVEP was enhanced when that color was attended versus when the other color was attended (figure 15.3), indicating again that attention can select specific color values independently of their particular location.

These demonstrations of global feature-based attention appear to conflict with the original finding of Hillyard and Münte (1984) that color selection was suppressed outside the focus of spatial attention. A critical difference in experimental design that may explain these disparate results, however, is that Hillyard and Münte presented stimuli briefly and intermittently, while all the studies that observed global feature selection presented the stimuli continuously. This difference suggests that the attended feature must actually be present in the display in order for global feature selection to override spatial selection.

Evidence that the relative priority of location- and feature-based selection may be flexibly adjusted according to task demands comes from ERP/MEG recordings in a visual search task (Hopf et al., 2004). Subjects searched for a simple color-orientation conjunction among distracters, half of which shared an orientation feature with the target and half of which did not (figure 15.4*A*). It was found that a lateralized brain response indicating the presence of the relevant orientation feature preceded the N2pc response indicating the spatial localization of the conjunction target by about 30 ms (figure 15.4*B*). These observations were taken to indicate that visual search involves a short phase of parallel, location-independent feature selection that occurs prior to the localization and selection of the target—a sequence of operations proposed by many influential theories of visual search (Cave, 1999; Treisman & Sato, 1990; Wolfe & Bennet, 1997).

Single-unit recording studies in monkeys (Bichot, Rossi, & Desimone, 2005) have provided additional evidence for parallel feature selection in visual search. Monkeys were

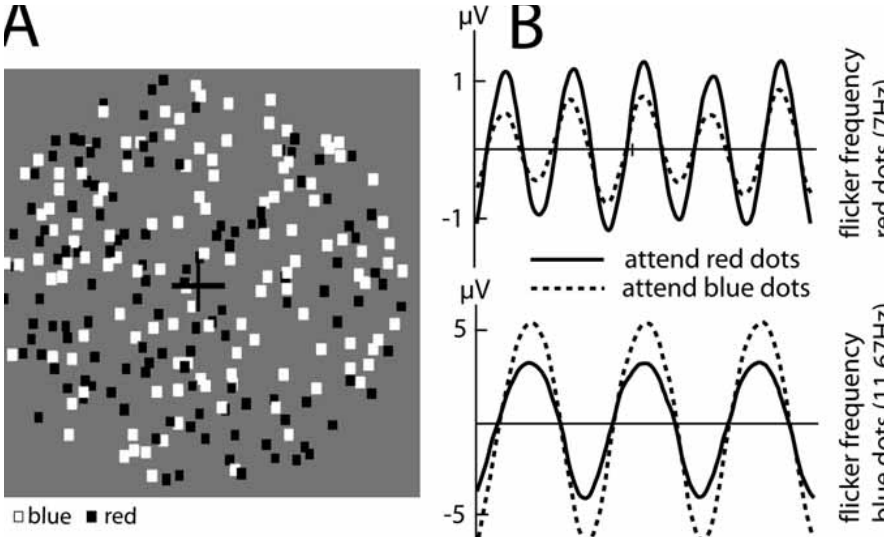


FIGURE 15.3 (*A*) Stimulus display in study by Müller et al. (2006), which consisted of 75 red dots (flickering at 7.0 Hz) intermingled with 75 blue dots (flickering at 11.67 Hz) within a 5 degree circle. Each dot changed its position in a random direction by 0.08 degrees every 1–3 frames of screen refresh (i.e., every 14.3–42.9 ms). On each trial (lasting 4.1 s) subjects were cued to attend to the red or blue dots and to discriminate occasional 586-ms periods of 75% coherent motion (targets) of the attended-color dots, which occurred on 32% of the trials. (*B*) SSVEPs from an occipital electrode elicited by the red and blue dot populations when attended and unattended, on trials without targets. Color-selective attention increased SSVEP amplitudes by an average of 30–40%. (Adapted from Müller et al., 2006.)

trained to search for a target defined by its color or form (or a combination of both) among many colored items of various forms. Search was performed under free gaze conditions, and the cell-firing responses in area V4 were analyzed during intermediate fixations along the search path toward the target, but before the target was fixated. It was observed that cell firing increased whenever a change in fixation brought a distracter that possessed one of the target's defining features into the cell's receptive field. These results showed that any item with a relevant feature was highlighted in area V4 even before ultimate target identification; this adds to the ERP/MEG evidence that attention to features acts in a "global" way prior to target selection in visual search.

A recent SSVEP study in humans (Andersen, Hillyard, & Müller, 2008) obtained evidence that the multiple features of an attended stimulus are selected and facilitated in a parallel, additive fashion. Subjects viewed a display containing 150 red and 150 blue bars, with half of each color oriented horizontally and half vertically, all of which were randomly intermixed and moving unpredictably. On each trial subjects were cued to attend to one of these four types of bars, each of which flickered at a different rate and thus elicited its own frequency-tagged SSVEP. It was found that SSVEP amplitudes were largest to the bars of the attended color orientation, intermediate to the bars having one of the two attended features, and smallest to the bars that lacked either attended feature. Most importantly, the SSVEP amplitude to the attended conjunction stimulus was equal to the sum of the amplitudes for the individual feature enhancements. This evidence for parallel, additive feature enhancement is consistent with the mechanism proposed by "guided search" theories (Wolfe et al., 1989) to account for the rapid identification of feature conjunction targets.

These physiological studies in monkeys and humans show that the time course and priority order of feature- and location-based attention effects can be flexibly adjusted depending on task demands and the particular selection operations that are required. The selection of nonspatial features is frequently delayed relative to the selection of locations but can be considerably accelerated for simple selections between feature dimensions. Moreover, feature selection may precede the selection of location in visual search tasks where the decoding of features is given priority to guide the subsequent spatial focusing of attention. In the next section we describe how such task-dependent flexibility also applies to the relation between feature- and object-based selection operations.

Object-based selection

A large body of psychophysical research indicates that attention can select entire objects for preferential processing (Driver, Davis, Russell, Turatto, & Freeman, 2001; Duncan & Nimmo-Smith, 1996; Egly, Driver, & Rafal, 1994; Scholl, 2001), which is not surprising given the ecological relevance of the multitudinous objects in the environment. It is important to realize, however, that object-based attention is a heterogeneous concept, and it is debated whether elemen-

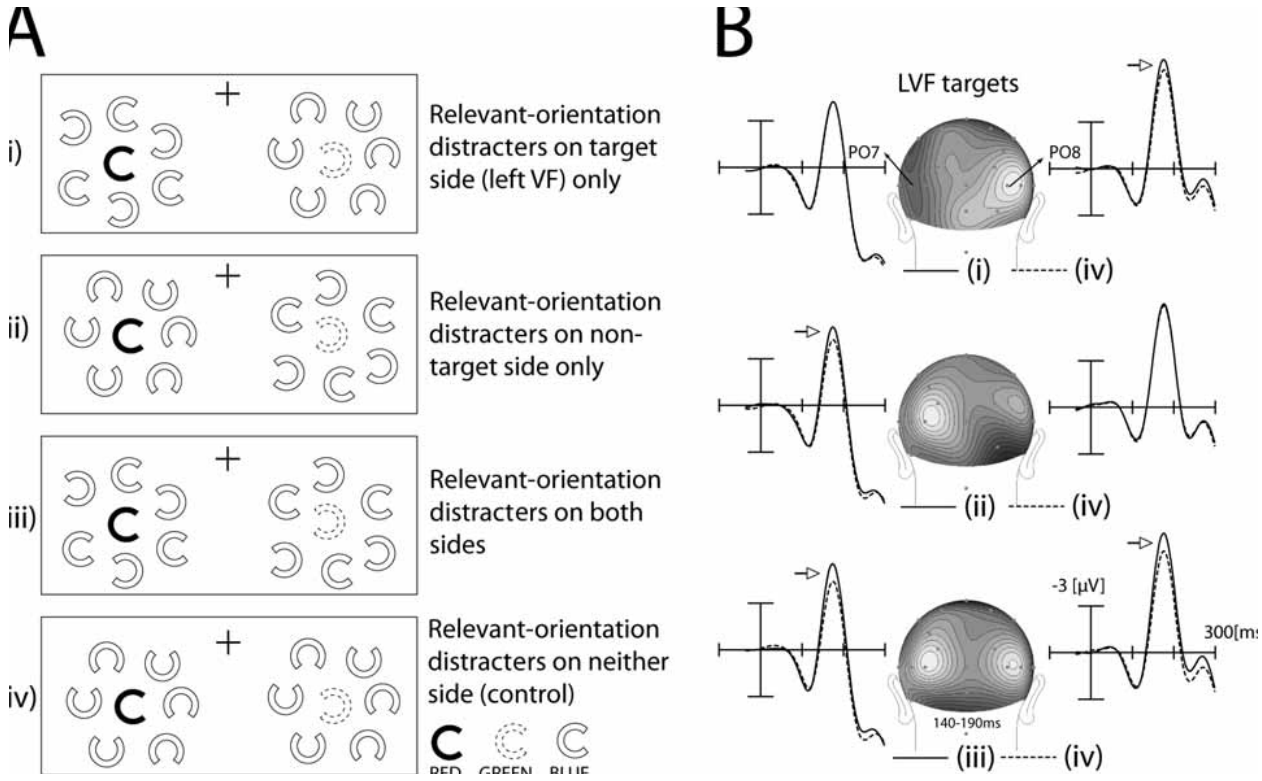


FIGURE 15.4 (*A*) Stimuli from the study of Hopf, Boelmans, Schoenfeld, Luck, and Heinze (2004). Search frames consisted of distinctively colored *C*'s (red and green, shown as black and dashed, respectively), one presented in the left and one in the right visual field, surrounded by blue distracter *C*s on each side. The red *C* served as target for half of the trial blocks and the green *C* for the other half. Subjects had to discriminate the orientation of the target *C* (here the red *C*, shown in black, in the left VF), whose gap always varied left-right. In contrast, distracters of one visual field were either oriented left-right, as was the target (relevant orientation distracters, RODs), or up-down (irrelevant orientation distracters). The location of RODs was varied relative to the location of the target item, such that RODs appeared (i) on the target side only, (ii) on the nontarget side only, (iii) on both sides, or (iv) on neither side (control condition). (*B*) ERP responses elicited by LVF targets. Waveforms of the different ROD-distributions (i–iii, solid tracings) are separately overlaid with the control condition (iv, dashed tracings). The topographical maps show the distribution of the corresponding voltage difference. The arrows highlight an enhanced negativity between ~140 and 300 ms that appears contralateral to the location of the RODs independent of the target's location in the left VF.

tary perceptual groupings (grouped array representations) or more abstract (spatially invariant) forms of representation serve as the objects of attention (Kramer, Weber, & Watson, 1997; Luck & Vecera, 2002; Mozer & Vecera, 2005; Vecera, 1997; Vecera & Farah, 1994; Weber, Kramer, & Miller, 1997). Another problem for isolating the neural mechanisms of object-based attention is that it appears far from trivial to rule out contributions from location- and feature-based mechanisms.

Valdes-Sosa and colleagues (Pinilla, Cobo, Torres, & Valdes-Sosa, 2001; Valdes-Sosa, Bobes, Rodriguez, & Pinilla, 1998) developed an elegant paradigm for studying object-based attention, which involves the competitive selection of perceived surfaces formed by moving dot arrays. The stimuli were two counterrotating random dot arrays that gave the impression of two superimposed surfaces rotating in opposite directions. This paradigm offers the opportunity to investigate the neural basis of object (surface) selection unconfounded by spatial attention, since the two rotating dot displays are superimposed. When attention was directed endogenously to one of the surfaces, observers could judge the direction of a brief translation of that surface and a second translation of the same surface much more accurately than translations of the uncued surface. Paralleling this perceptual selection, ERP recordings showed that occipital P1 and N1 (N200) components elicited by translations of the unattended surface were suppressed relative to those of the attended surface. This finding was taken to indicate that attention favors processing of the attended surface by attenuating the object representation of the unattended surface. A subsequent study (Rodriguez & Valdes-Sosa, 2006) carried out current source localization of the ERP modulation reflecting this motion-based surface selection and found that the associated N200 component was generated in part in human MT+, which corresponds to recent findings from a neurophysiological study of surface

segmentation in monkey MT (Wannig, Rodriguez, & Freiwald, 2007).

Reynolds, Alborzian, and Stoner (2003) replicated the findings of Valdes-Sosa's group and showed in addition that a brief translation of one of the surfaces was an effective exogenous cue that automatically engaged attention to that surface, resulting in superior detection of a second translation of the same surface. In an ERP study of this paradigm, Khoe, Mitchell, Reynolds, and Hillyard (2005) found that when the first and second translations were of the same surface (relative to different surfaces), the second translation was discriminated more accurately and elicited an N1 component of greater amplitude. It was concluded that object-based surface selection can be initiated by purely exogenous cuing, resulting in a stronger representation of the cued surface in the visual pathways. A further study (Mitchell, Stoner, & Reynolds, 2004) showed that cuing (translating) one of the rotating surfaces resulted in a sustained dominance for the eye viewing that cued surface when the surfaces were switched to dichoptic viewing (one surface to each eye) immediately after the cue. This binocular rivalry/ocular dominance effect was presumably mediated by the object-based selection of the surface, which was conveyed to a neural representation with eye-of-origin information in visual cortex. This interpretation was confirmed in an ERP study of this paradigm (Khoe, Mitchell, Reynolds, & Hillyard, 2008). As in the original study, subjects were impaired at comparing the first and second translations when they belonged to different surfaces, and this impairment was greater when the surfaces were viewed dichoptically (in rivalry). The P1 component (110–160 ms) in the ERP elicited by the second translation of the same surface was larger than in the ERP elicited by the different surface during dichoptic but not monocular viewing conditions. A larger cuing effect was also found for the subsequent N1 (160–220 ms) during rivalry than during monocular viewing. It was concluded that surface selection can occur at an earlier level of processing when the cued and uncued surfaces are presented to the separate eyes (for similar results see Mishra & Hillyard, 2008).

An important line of evidence that attention operates in an object-based framework comes from the seminal paradigm developed by Egly and colleagues (1994). This design is an extension of Posner's classical cuing paradigm, in which a comparison is made between validly and invalidly cued locations that belong either to the same or to separate objects. A typical observation indicating the operation of object-based attention has been that performance costs at an invalidly cued location were significantly reduced when this location belonged to the cued object versus to the uncued object. Importantly, the Egly type of paradigm provides a framework for investigating the relationship between space-based and object-based attention.

Recent ERP studies using variations of the Egly paradigm (He, Fan, Zhou, & Chen, 2004; Martinez, Ramanathan, Foxe, Javitt, & Hillyard, 2007; Martinez, Teder-Salejarvi, & Hillyard, 2007; Martinez et al., 2006) have revealed that modulations due to object-based selection have important similarities to those seen during space-based selection. For example, in a combined ERP/fMRI study Martinez and colleagues (2006) presented subjects with a display consisting of two bar-shaped objects oriented either horizontally or vertically (figure 15.5) while continuous sequences of stimuli (brief corner offsets) were presented one at a time in random order at both ends of both bars. On each run the corner offsets at one end of one bar were cued to be attended while stimuli at the other three locations were ignored. It was found (figure 15.6) that unattended corner offsets produced a larger amplitude N1 component (but not the P1 component) when they belonged to the attended bar versus when they belonged to the unattended bar (see He et al., 2004, for a similar observation). Although this object-mediated N1 effect (at 160–200 ms) was smaller than the N1 enhancement to attended-location stimuli in the same latency range, both effects showed very similar posterior-contralateral scalp distributions. Furthermore, electric source analysis (confirmed by fMRI in parallel experiments) revealed that both the object-based and location-based N1 enhancements were generated in the same area of the lateral occipital cortex (LOC)—a cortical region that has been strongly implicated as being critically involved in the initial segmentation and encoding of objects (e.g., Grill-Spector, 2003). Similar effects were obtained even when using objects defined by illusory contours (Martinez, Teder-Salejarvi, et al., 2007) and objects of different shapes (Martinez, Ramanathan, et al., 2007). These ERP experiments and a recent fMRI study (Müller & Kleinschmidt, 2003) provide evidence that object-selective attention shares a common neural mechanism with spatial attention that facilitates the sensory processing of stimuli within the boundaries of an attended object. These findings support the view that spatial attention directed to one part of an object spreads throughout its boundaries and strengthens the sensory representation of the entire object (Vecera & Farah, 1994; Weber et al., 1997). This selective enhancement of attended object representations, particularly in area LOC, may reinforce their perceptual integrality and underlie the performance benefits manifested in the same-object advantage.

A major question regarding the neural mechanisms of object-based attention is how the different features of an object, which may be represented in widely dispersed cortical areas, are bound together to form a unified percept. One approach to this "binding problem" comes from Duncan's "integrated competition model" (Duncan, Humphreys, & Ward, 1997). According to this model, directing attention to one of an object's features produces a competitive advantage for the object in the neural module encoding that feature,

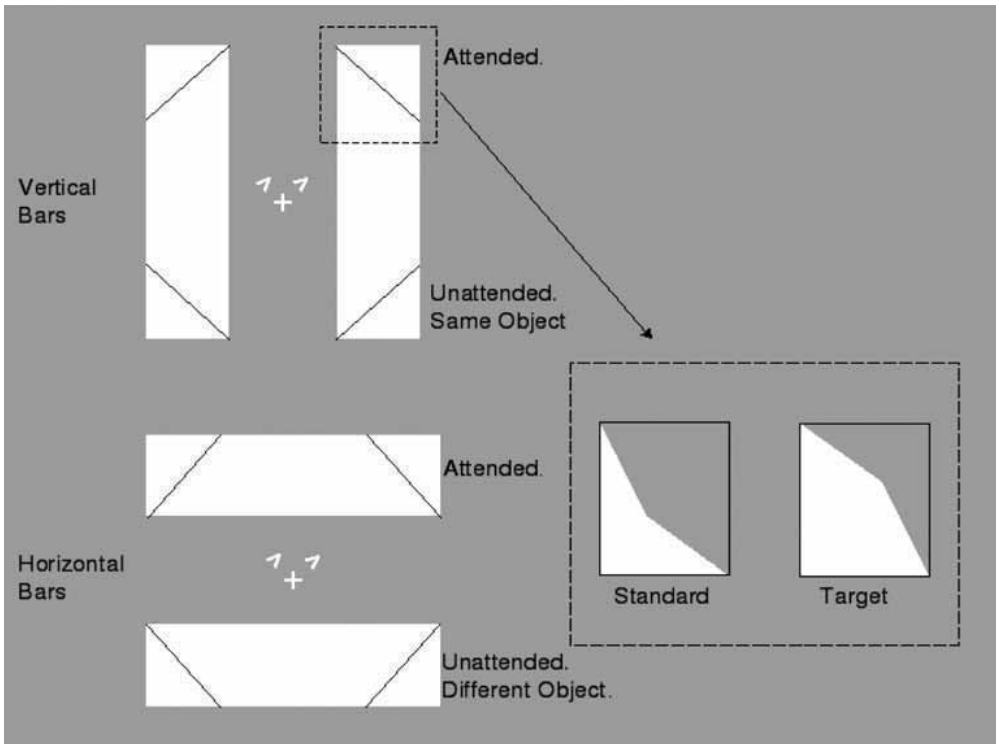


FIGURE 15.5 Experimental design in study by Martinez and colleagues (2006). During each run, either two horizontal or two vertical bars were presented continuously on the screen. Subjects were cued by a pair of arrows near fixation to attend covertly to one of the four visual quadrants. Stimuli were brief (100 ms) offsets of the corners of the bars, leaving either a concave (standard) or convex (target) edge. Corner offsets occurred in random order in all four quadrants with ISIs of 400–600 ms. Subjects responded to detections of targets in the attended quadrant, which occurred with a probability of 0.2. In the example shown, the upper left quadrant was attended; thus the unattended lower left quadrant belonged to the attended object when the bars were vertical but not when the bars were horizontal.

which is then transmitted to the modules encoding the other features of the object. The resulting activation of the entire network of specialized modules underlies the binding of features into an integrated perceptual object. A key prediction of the integrated competition model is that directing attention to one feature of an object will result in the activation of its other features, including those that are irrelevant to the task at hand. Such an effect was demonstrated in an fMRI study by O'Craven, Downing, & Kanwisher (1999), who presented subjects with superimposed transparent pictures of houses and faces, with one moving and the other being stationary. On different runs subjects selectively attended to houses, to faces, or to stimulus motion itself. It was found that neural activity was increased not only in the critical area specific to the attended stimulus feature (the fusiform face area for faces; the parahippocampal place area for houses; area MT+ for motion), but also in the area encoding the task-irrelevant feature of the attended object. These data provided strong support for object-based selection in the context of the integrated competition model.

While the study of O'Craven and colleagues was elegantly designed, it remains possible that location cues played some role, as the objects were not perfectly overlapping and target-defining details might have been spatially selected. To verify a mechanism of object-based selection, it is important to eliminate potential contributions from other possible reference frames. In this respect, a convincing behavioral demonstration of object-based attention that eliminated all possible alternative explanations was carried out by Blaser, Pylyshyn, & Holcombe (2000). Subjects had to track the identity of one of two spatially superimposed Gabor patches, which continuously changed along feature dimensions of orientation, color, and spatial frequency. This approach not only eliminated the possibility of location-based selection, but also ruled out featural identity as a basis for tracking the target. The observation that subjects were still able to track the identity of the Gabor patch over time confirmed that visual attention had been directed to an object-based representation.

The key finding of O'Craven and colleagues (1999) was that directing attention to one feature of an object activated the neural representations of its other features, including those not relevant to the current task. Given the low time resolution of fMRI, however, it was not possible to determine whether the activation of irrelevant features occurs rapidly enough to participate in the feature-binding process that underlies the perception of the integrated object. This

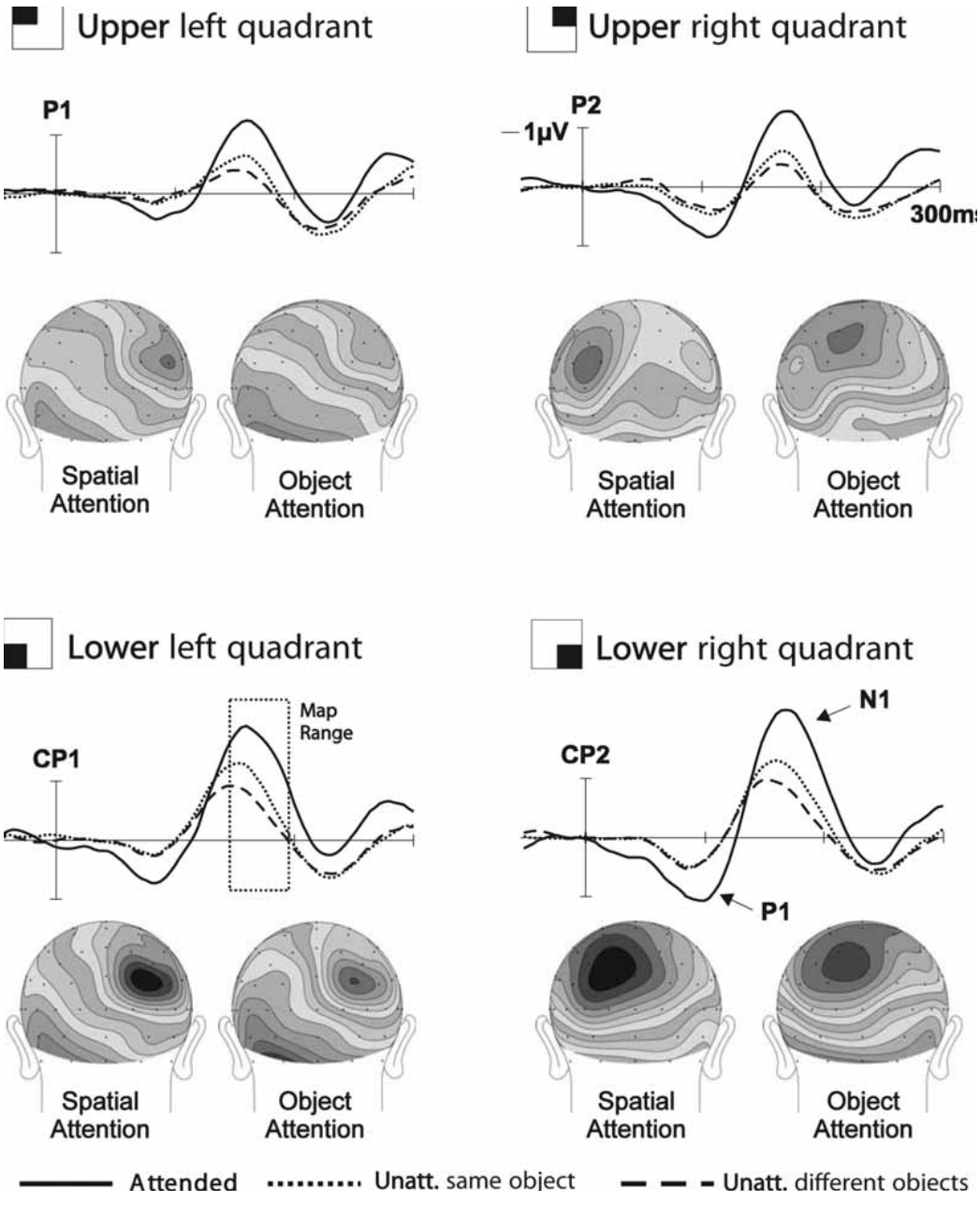


FIGURE 15.6 Grand average ERPs to nontarget stimuli in the four quadrants in the study of Martinez and colleagues (2006). Overlapped waveforms are ERPs to stimuli when attended (solid lines), when unattended but belonging to the attended object (dotted lines), and when unattended and belonging to the unattended object (dashed lines). For each quadrant, voltage maps of attention related difference waves in the latency range of the N1 (160–196 ms) are shown: the spatial attention difference was obtained by subtracting the unattended from unattended amplitudes; the object attention difference was obtained by subtracting the unattended-location/unattended-object amplitude from the unattended-location/attended-object amplitude. Note the similarity between the object and spatial attention topographies. Increasing darkness on maps indicates increased negativity. Contour intervals are 0.05 μV for the object attention maps and 0.15 μV for the spatial attention maps.

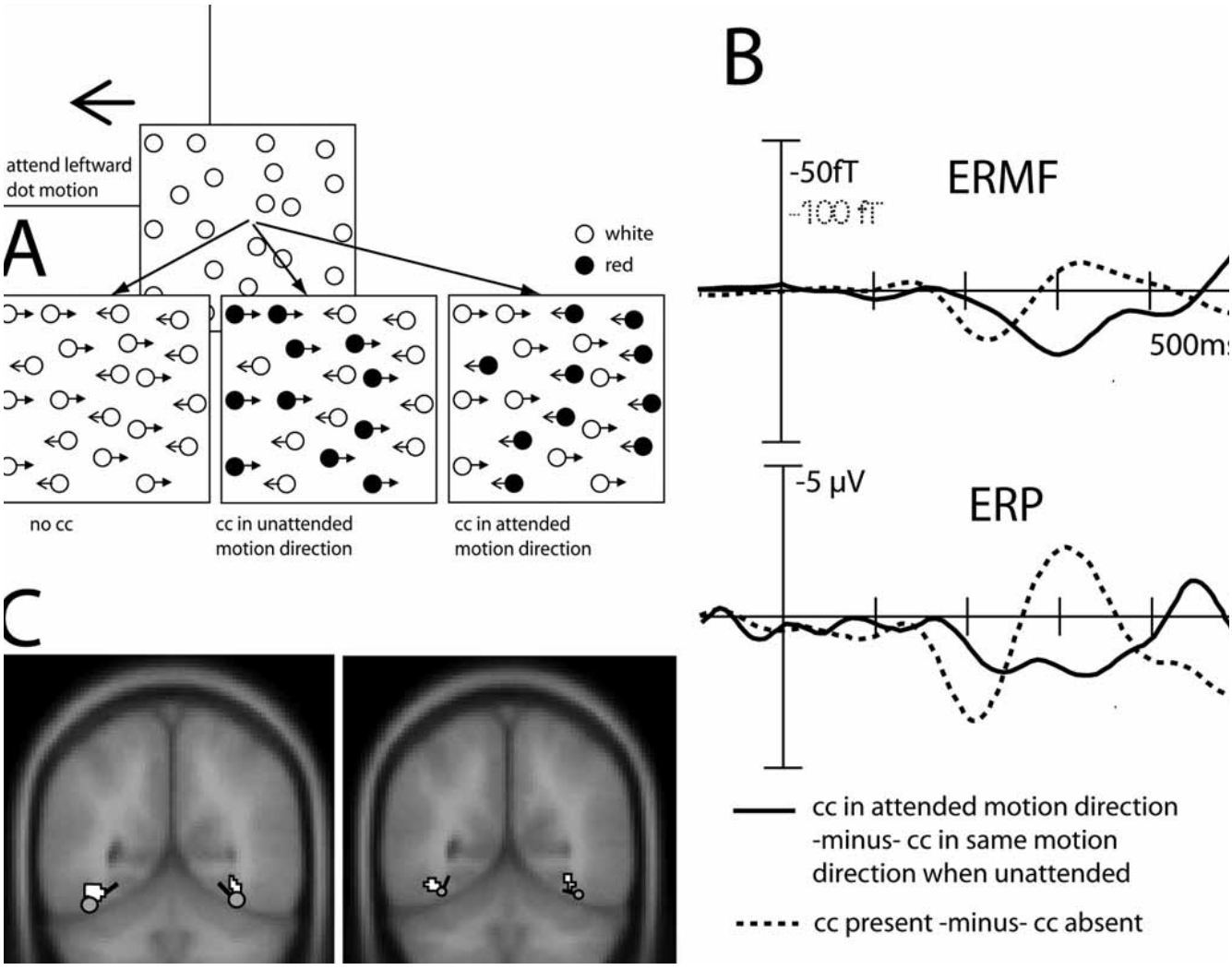


FIGURE 15.7 (*A*) Experimental design of study by Schoenfeld and colleagues (2003). On each trial a random half of the dots on the screen moved left and half right for 300 ms. Subjects were cued to attend to either the left- or right-moving dots and reported occasional targets of higher velocity. On a random basis, either the left- or right-moving dots could change color (CC) to red or remain white. (*B*) The sensory effect of color is shown in the dotted ERMF and ERP waveforms: these are difference waves formed by subtracting responses to the no CC condition from the condition where the CC belonged to the unattended dots. The attention-related activation of the irrelevant color feature is represented by the solid waveforms, formed by subtracting responses when the CC belonged to the unattended dots minus when the CC belonged to the attended dots. Note that the attention-produced enhancement of the irrelevant color feature lags the initial sensory response to color by 30–50 ms. (*C*) The neural generators of this increased color signal at 220–300 ms were localized to the ventral occipital cortex (area V4v), known to be a specialized module for color processing (McKeefry & Zeki, 1997). This localization was confirmed by BOLD signal activations (irregular areas on brain slices) in a parallel experiment using the same design with fMRI.

question was investigated in a study that combined recordings of ERPs and MEG with fMRI in a task where subjects attended to multifeature objects formed by arrays of dots on a video screen (Schoenfeld et al., 2003). On each trial half of the dots on the screen moved to the left and half to the right, producing the perception of two transparent surfaces moving in opposite directions (figure 15.7*A*). On each trial subjects were cued to attend to either the left-moving or the right-moving surface, and on a random basis one of the surfaces underwent an irrelevant color change (white to red). When physically identical trials with the attended versus unattended surface changing color were compared, increased fMRI activation was found in the color-specific region of the fusiform gyrus area V4v. More importantly, the ERP and MEG recordings showed enhanced amplitudes at 220–240 ms after stimulus onset on trials when the attended surface was colored (figure 15.7*B*), and the neural generators of this activity were localized to the same V4v site (figure 15.7*C*). This attention-related amplification of the color signal occurred about 40–50 ms after the initial arrival of color-specific input to area V4. It was concluded that this amplification of an irrelevant feature of an attended object occurs rapidly enough to participate in the feature-binding process that underlies the perceptual unity of the attended object. These results fit nicely with Duncan's (Duncan et al., 1997) integrated competition model and provide specific information about the time scale of its operation.

An observation complementary to that of Schoenfeld and colleagues (2003) was recently made with single-unit recordings in macaque V4 (Fallah, Stoner, & Reynolds, 2007). In this study, superimposed rotating dot surfaces of different color were used. To bias the monkey to attend one of the surfaces, the motion onset of one surface was delayed relative to the other surface, thereby producing a predictable bias for the delayed surface. Firing activity was then recorded from V4 cells whose RF overlapped with the dot surfaces. Enhanced firing was seen when the delayed-onset surface was of the neuron's preferred color, while a relative suppression appeared when the delayed-onset color was of the nonpreferred color. Both effects appeared rapidly, with a delay of 10–17 ms relative to the activity elicited by motion onset of a single surface with the preferred color. This finding, together with the results of Schoenfeld and colleagues (2003), points to a fast but clearly delayed facilitation of task-irrelevant features bound into the attended object, which emphasizes the rapid evolution of object-based surface segmentation under those experimental conditions.

Conclusion

The studies reviewed in this chapter summarize the spatiotemporal signatures of space-, feature- and object-based attention as revealed by electromagnetic and electrophysiological recordings, as well as by functional brain imaging. All three aspects of attention turn out to involve multiple selection operations, which are coordinated on a tight temporal scale of a few tens of milliseconds in early visual processing. We have seen that the temporal priority of those selection operations may change flexibly depending on the particular requirements of visual decoding, task demands, and experimental instructions. The selection of features may precede the selection of locations or objects in visual search, but object-based selection may take priority under other conditions. Moreover, the neural mechanisms that underlie each of the different forms of attention may be deployed in a flexible manner. For example, spatial attention involves biasing mechanisms that can be adjusted to select stimuli at single or multiple locations, or at even more complex spatial arrays. Finally, by highlighting the ways in which space-, feature- and object-based attention interact, it has become evident that there may not be a strict demarcation line. Space-based attention was found to facilitate object-based attention, which in turn promoted the selection of the object's features. Given the enormity of the task of controlling the order of information flow in human vision, this multiplicity, flexibility, and interdependence of attentional operations does not come as a surprise.

REFERENCES

Aine, C. J., Supek, S., & George, J. S. (1995). Temporal dynamics of visual-evoked neuromagnetic sources: Effects of stimulus parameters and selective attention. *Int. J. Neurosci.*, *80*, 79–104.

Andersen, S. K., Hillyard, S. A., & Müller, M. M. (2008). Attention facilitates multiple stimulus features in parallel in human visual cortex. *Curr. Biol*, *18*(13), 1006–1009.

Anllo-Vento, L., & Hillyard, S. A. (1996). Selective attention to the color and direction of moving stimuli: Electrophysiological correlates of hierarchical feature selection. *Percept. Psychophys.*, *58*, 191–206.

Anllo-Vento, L., Luck, S. J., & Hillyard, S. A. (1998). Spatio-temporal dynamics of attention to color: Evidence from human electrophysiology. *Hum. Brain Mapp.*, *6*, 216–238.

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *J. Exp. Psychol. Hum. Percept. Perform.*, *26*, 834–846.

Baas, J. M., Kenemans, J. L., & Mangun, G. R. (2002). Selective attention to spatial frequency: An ERP and source localization analysis. *Clin. Neurophysiol.*, *113*, 1840–1854.

Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vis. Res.*, *39*, 71–86.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*, 529–534.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, *408*, 196–199.

Bullier, J. (2001a). Feedback connections and conscious vision. *Trends Cogn. Sci.*, *5*, 369–370.

Bullier, J. (2001b). Integrated model of visual processing. *Brain Res. Brain Res. Rev.*, *36*, 96–107.

Bundesen, C. (1990). A theory of visual attention. *Psychol. Rev.*, *97*, 523–547.

Castiello, U., & Umilta, C. (1992). Splitting focal attention. *J. Exp. Psychol. Hum. Percept. Perform.*, *18*, 837–848.

Cave, K. R. (1999). The FeatureGate model of visual selection. *Psychol. Res.*, *62*, 182–194.

Cave, K. R., & Pashler, H. (1995). Visual selection mediated by location: Selecting successive visual objects. *Percept. Psychophys.*, *57*, 421–432.

Cave, K. R., & Zimmerman, J. M. (1997). Flexibility in spatial attention before and after practice. *Psychol. Sci.*, *8*, 399–403.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nat. Neurosci.*, *2*, 671–676.

Clark, V. P., Fan, S., & Hillyard, S. A. (1995). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Hum. Brain Mapp.*, *2*, 170–187.

Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *J. Cogn. Neurosci.*, *8*, 387–402.

Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *J. Neurophysiol.*, *75*, 1306–1308.

Connor, C. E., Preddy, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *J. Neurosci.*, *19*, 3201–3214.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, *248*, 1556–1559.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: Functional anatomy by positron emission tomography. *J. Neurosci.*, *11*, 2383–2402.

Cutzu, F., & Tsotsos, J. K. (2003). The selective tuning model of attention: Psychophysical evidence for a suppressive annulus around an attended item. *Vis. Res.*, *43*, 205–219.

Di Russo, F., Martinez, A., & Hillyard, S. A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cereb. Cortex*, *13*, 486–499.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Hum. Brain Mapp.*, *15*, 95–111.

Downing, P. E., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner, & O. S. Marin (Eds.), *Attention and Performance XI* (pp. 171–188). Hillsdale, NJ: Erlbaum.

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition*, *80*, 61–95.

Duncan, J., Humphreys, G. W., & Ward, R. M. (1997). Competitive brain activity in visual attention. *Curr. Opin. Neurobiol.*, *7*, 255–261.

Duncan, J., & Nimmo-Smith, I. (1996). Objects and attributes in divided attention: Surface and boundary systems. *Percept. Psychophys.*, *58*, 1076–1084.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *J. Exp. Psychol. Gen.*, *123*, 161–177.

Eimer, M. (1997a). Attentional selection and attentional gradients: An alternative method for studying transient visual-spatial attention. *Psychophysiology*, *34*, 365–376.

Eimer, M. (1997b). An event-related potential (ERP) study of transient and sustained visual attention to color and form. *Biol. Psychol.*, *44*, 143–160.

Eimer, M. (1999). Attending to quadrants and ring-shaped regions: ERP effects of visual attention in different spatial selection tasks. *Psychophysiology*, *36*, 491–503.

Eimer, M. (2000). An ERP study of sustained spatial attention to stimulus eccentricity. *Biol. Psychol.*, *52*, 205–220.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Percept. Psychophys.*, *12*, 201–204.

Eriksen, C. W., & Yeh, Y.-Y. (1985). Allocation of attention in the visual field. *J. Exp. Psychol. Hum. Percept. Perform.*, *11*, 583–597.

Fallah, M., Stoner, G. R., & Reynolds, J. H. (2007). Stimulus-specific competitive selection in macaque extrastriate visual area V4. *Proc. Natl. Acad. Sci. USA*, *104*, 4165–4169.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining "early" visual processing. *Exp. Brain Res.*, *142*, 139–150.

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*, 677–696.

Grill-Spector, K. (2003). The neural basis of object perception. *Curr. Opin. Neurobiol.*, *13*, 159–166.

Hahn, S., & Kramer, A. F. (1998). Further evidence for the division of attention between noncontiguous locations. *Visual Cogn.*, *5*, 217–256.

Handy, T. C., Kingstone, A., & Mangun, G. R. (1996). Spatial distribution of visual attention: Perceptual sensitivity and response latency. *Percept. Psychophys.*, *58*, 613–627.

Handy, T. C., & Mangun, G. R. (1997). *Early attention selection: Electrophysiological evidence for modulation by perceptual load.* Manuscript.

Handy, T. C., & Mangun, G. R. (2000). Attention and spatial selection: Electrophysiological evidence for modulation by perceptual load. *Percept. Psychophys.*, *62*, 175–185.

Harter, M. R., & Aine, C. (1984). Brain mechanisms of visual selective attention. In R. Parasuraman, & D. R. Davies (Eds.), *Varieties of attention* (pp. 293–321). New York: Academic Press.

He, X., Fan, S., Zhou, K., & Chen, L. (2004). Cue validity and object-based attention. *J. Cogn. Neurosci.*, *16*, 1085–1097.

Heinze, H. J., Luck, S. J., Münte, T. F., GÖs, A., Mangun, G. R., & Hillyard, S. A. (1994). Attention to adjacent and separate positions in space: An electrophysiological analysis. *Percept. Psychophys.*, *56*, 42–52.

Heinze, H. J., Mangun, G. R., Burchert, W., Hinrichs, H., Scholz, M., Münte, T. F., et al. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature*, *372*, 543–546.

Helmholtz, H. V. (1896). *Handbuch der physiologischen Optik.* Hamburg: Verlag von Leopold Voss.

Henderson, J. M., & MacQuistan, A. D. (1993). The spatial distribution of attention following an exogenous cue. *Percept. Psychophys.*, *53*, 221–230.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Natl. Acad. Sci. USA*, *95*, 781–787.

Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Percept. Psychophys.*, *36*, 185–198.

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. London B Biol. Sci.*, *353*, 1257–1270.

Hopf, J. M., Boehler, N., Luck, S. J., Tsotsos, J. K., Heinze, H.-J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proc. Natl. Acad. Sci. USA*, *103*, 1053–1058.

Hopf, J. M., Boelmans, K., Schoenfeld, A., Luck, S. J., & Heinze, H.-J. (2004). Attention to features precedes attention to locations in visual search: Evidence from electromagnetic brain responses in humans. *J. Neurosci.*, *24*, 1822–1832.

Hopf, J.-M., Luck, S. J., Girelli, M., Hagner, T., Mangun, G. R., Scheich, H., et al. (2000). Neural sources of focused attention in visual search. *Cereb. Cortex*, *10*, 1233–1241.

Hopf, J. M., Schoenfeld, M. A., & Heinze, H. J. (2005). The temporal flexibility of attentional selection in the visual cortex. *Curr. Opin. Neurobiol.*, *15*, 183–187.

Hopf, J. M., Vogel, E., Woodman, G., Heinze, H. J., & Luck, S. J. (2002). Localizing visual discrimination processes in time and space. *J. Neurophysiol.*, *88*, 2088–2095.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat. Neurosci.*, *3*, 284–291.

Hopfinger, J. B., Luck, S. J., & Hillyard, S. (2004). Selective attention: Electrophysiological and neuromagnetic studies. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (3rd ed., pp. 561–574). Cambridge, MA: Bradford, MIT Press.

Hopfinger, J. B., & Mangun, G. R. (1998). Reflexive attention modulates processing of visual stimuli in human extrastriate cortex. *Psychol. Sci.*, *9*, 441–447.

Hopfinger, J. B., & Mangun, G. R. (2001). Tracking the influence of reflexive attention on sensory and cognitive processing. *Cogn. Affective Behav. Neurosci.*, *1*, 56–65.

Hopfinger, J. B., & Ries, A. J. (2005). Automatic versus contingent mechanisms of sensory-driven neural biasing and reflexive attention. *J. Cogn. Neurosci.*, *17*, 1341–1352.

Juola, J. F., Bouwhuis, D. G., Cooper, E. E., & Warner, B. (1991). Control of attention around the fovea. *J. Exp. Psychol. Hum. Percept. Perform.*, *17*, 125–141.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*, 751–761.

Kelly, S. P., Gomez-Ramirez, M., & Foxe, J. J. (2008). Spatial attention modulates initial afferent activity in human primary visual cortex. *Cereb. Cortex*, *18*, 2629–2636.

Kenemans, J. L., Lijffijt, M., Camfferman, G., & Verbaten, M. N. (2002). Split-second sequential selective activation in human secondary visual cortex. *J. Cogn. Neurosci.*, *14*, 48–61.

Khoe, W., Mitchell, J. F., Reynolds, J. H., & Hillyard, S. A. (2005). Exogenous attentional selection of transparent superimposed surfaces modulates early event-related potentials. *Vis. Res.*, *45*, 3004–3014.

Khoe, W., Mitchell, J. F., Reynolds, J. H., & Hillyard, S. A. (2008). ERP evidence that surface-based attention biases interocular competition during rivalry. *J. Vis.*, *8*(3):18, 1–11.

Kramer, A. F., & Hahn, S. (1995). Splitting the beam: distribution of attention over noncontiguous regions of the visual field. *Psychol. Sci.*, *6*, 381–386.

Kramer, A. F., Weber, T. A., & Watson, S. E. (1997). Object-based attentional selection—Grouped arrays or spatially invariant representations? Comment on Vecera and Farah (1994). *J. Exp. Psychol. Gen.*, *126*, 3–13.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.*, *23*, 571–579.

Lee, J., Williford, T., & Maunsell, J. H. (2007). Spatial attention and the latency of neuronal responses in macaque area V4. *J. Neurosci.*, *27*, 9632–9637.

Liu, T., Slotnick, S. D., Serences, J. T., & Yantis, S. (2003). Cortical mechanisms of feature-based attentional control. *Cereb. Cortex*, *13*, 1334–1343.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.*, *77*, 24–42.

Luck, S. J., Girelli, M., Mc^{d}ermott, M. T., & Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cogn. Psychol.*, *33*, 64–87.

Luck, S. J., & Hillyard, S. A. (1994a). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*, 291–308.

Luck, S. J., & Hillyard, S. A. (1994b). Spatial filtering during visual search: Evidence from human electrophysiology. *J. Exp. Psychol. Hum. Percept. Perform.*, *20*, 1000–1014.

Luck, S. J., & Hillyard, S. A. (1999). The operation of selective attention at multiple stages of processing: Evidence from human and monkey electrophysiology. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 687–700). Cambridge, MA: MIT Press.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cueing on luminance detectability: Psychophysical and electrophysical evidence for early selection. *J. Exp. Psychol. Hum. Percept. Perform.*, *20*, 887–904.

Luck, S. J., & Vecera, S. P. (2002). Attention: From paradigms to mechanisms. In S. Yantis (Ed.), *Stevens' handbook of experimental psychology: Sensation and perception* (pp. 235–286). New York: Wiley.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cogn. Sci.*, *4*, 432–440.

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*, 4–18.

Mangun, G. R., Buonocore, M., Girelli, M., & Jha, A. (1998). ERP and fMRI measures of visual spatial selective attention. *Hum. Brain Mapp.*, *6*, 383–389.

Mangun, G. R., & Hillyard, S. A. (1990). Allocation of visual attention to spatial locations: Tradeoff functions for event-related brain potentials and detection performance. *Percept. Psychophys.*, *47*, 532–550.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked potentials indicate changes in perceptual processing during visual-spatial priming. *J. Exp. Psychol. Hum. Percept. Perform.*, *17*, 1057–1074.

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci*, *2*, 364–369.

Martinez, A., Di Russo, F., Anllo-Vento, L., & Hillyard, S. A. (2001). Electrophysiological analysis of cortical mechanisms of selective attention to high and low spatial frequencies. *Clin. Neurophysiol.*, *112*, 1980–1998.

Martinez, A., Di Russo, F., Anllo-Vento, L., Sereno, M. I., Buxton, R. B., & Hillyard, S. A. (2001). Putting spatial attention on the map: Timing and localization of stimulus

selection processes in striate and extrastriate visual areas. *Vis. Res.*, *41*, 1437–1457.

Martinez, A., Ramanathan, D. S., Foxe, J. J., Javitt, D. C., & Hillyard, S. A. (2007). The role of spatial attention in the selection of real and illusory objects. *J. Neurosci.*, *27*, 7963–7973.

Martinez, A., Teder-Salejarvi, W., & Hillyard, S. A. (2007). Spatial attention facilitates selection of illusory objects: Evidence from event-related brain potentials. *Brain Res.*, *1139*, 143–152.

Martinez, A., Teder-Salejarvi, W., Vazquez, M., Molholm, S., Foxe, J. J., Javitt, D. C., et al. (2006). Objects are highlighted by spatial attention. *J. Cogn. Neurosci.*, *18*, 298–310.

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.*, *14*, 744–751.

Maunsell, J. H., & Cook, E. P. (2002). The role of attention in visual processing. *Philos. Trans. R. Soc. Lon. B Biol. Sci.*, *357*, 1063–1072.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci.*, *29*, 317–322.

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.*, *19*, 431–441.

McKeefry, D. J., & Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*, *120*, 2229–2242.

McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron*, *42*, 677–686.

McMains, S. A., & Somers, D. C. (2005). Processing efficiency of divided spatial attention mechanisms in human visual cortex. *J. Neurosci.*, *25*, 9444–9448.

Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2000a). Intermodal selective attention in monkeys. I. Distribution and timing of effects across visual areas. *Cereb. Cortex*, *10*, 343–358.

Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2000b). Intermodal selective attention in monkeys. II. Physiological mechanisms of modulation. *Cereb. Cortex*, *10*, 359–370.

Mishra, J., & Hillyard, S. A. (2008). Endogenous attention selection during binocular rivalry at early stages of visual processing. *Vis. Res.* [Epub ahead of print. doi: 10.1016/j.visres.2008.02.018.]

Mitchell, J. F., Stoner, G. R., & Reynolds, J. H. (2004). Object-based attention determines dominance in binocular rivalry. *Nature*, *429*, 410–413.

Motter, B. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.*, *70*, 909–919.

Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.*, *14*, 2178–2189.

Mounts, J. R. (2000). Attentional capture by abrupt onsets and feature singletons produces inhibitory surrounds. *Percept. Psychophys.*, *62*, 1485–1493.

Mozer, M. C., & Vecera, S. P. (2005). Object- and space-based attention. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 130–134). Burlington, MA: Elsevier/Academic Press.

Müller, M. M., Andersen, S., Trujillo, N. J., Valdes-Sosa, P., Malinowski, P., & Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proc. Natl. Acad. Sci. USA*, *103*, 14250–14254.

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, *424*, 309–312.

Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A physiological correlate of the "zoom lens" of visual attention. *J. Neurosci.*, *23*, 3561–3565.

Müller, N. G., & Kleinschmidt, A. (2003). Dynamic interaction of object- and space-based attention in retinotopic visual areas. *J. Neurosci.*, *23*, 9812–9816.

Müller, N. G., & Kleinschmidt, A. (2004). The attentional "spotlight's" penumbra: Center-surround modulation in striate cortex. *NeuroReport*, *15*, 977–980.

Müller, N. G., Mollenhauer, M., Rüsler, A., & Kleinschmidt, A. (2005). The attentional field has a Mexican hat distribution. *Vis. Res.*, *45*, 1129–1137.

Noesselt, T., Hillyard, S., Woldorff, M., Schoenfeld, A., Hagner, T., Jancke, L., et al. (2002). Delayed striate cortical activation during spatial attention. *Neuron*, *35*, 575–587.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, *18*, 591–598.

Olson, I. R., Chun, M. M., & Allison, T. (2001). Contextual guidance of attention: Human intracranial event-related potential evidence for feedback modulation in anatomically early temporally late stages of visual processing. *Brain*, *124*, 1417–1425.

Pinilla, T., Cobo, A., Torres, K., & Valdes-Sosa, M. (2001). Attentional shifts between surfaces: Effects on detection and early brain potentials. *Vis. Res.*, *41*, 1619–1630.

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.*, *32*, 3–25.

Reynolds, J. H., Alborzian, S., & Stoner, G. R. (2003). Exogenously cued attention triggers competitive selection of surfaces. *Vis. Res.*, *43*, 59–66.

Rodriguez, V., & Valdes-Sosa, M. (2006). Sensory suppression during shifts of attention between surfaces in transparent motion. *Brain Res.*, *1072*, 110–118.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of macaque monkey. *Nature*, *395*, 376–381.

Roelfsema, P. R., Tolboom, M., & Khayat, P. S. (2007). Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron*, *56*, 785–792.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.*, *5*, 631–632.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vis. Res.*, *43*, 629–637.

Schoenfeld, M., Hopf, J. M., Martinez, A., Mai, H., Sattler, C., Gasde, A., et al. (2007). Spatio-temporal analysis of feature-based attention. *Cereb. Cortex*, *17*, 2468–2477.

Schoenfeld, M. A., Tempelmann, C., Martinez, A., Hopf, J. M., Sattler, C., Heinze, H. J., et al. (2003). Dynamics of feature binding during object- selective attention. *Proc. Natl. Acad. Sci. USA*, *100*, 11806–11811.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*, 1–46.

Serences, J. T., Yantis, S., Culberson, A., & Awh, E. (2004). Preparatory activity in visual cortex indexes distractor suppression during covert spatial orienting. *J. Neurophysiol.*, *92*, 3538–3545.

Shipp, S. (2007). Structure and function of the cerebral cortex. *Curr. Biol.*, *17*, R443–449.

Shulman, G. L., Wilson, J., & Sheehy, J. B. (1985). Spatial determinants of the distribution of attention. *Percept. Psychophys.*, *37*, 59–65.

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *J. Neurophysiol.*, *97*, 229–237.

Slotnick, S. D., Hopfinger, J. B., Klein, S. A., & Sutter, E. E. (2002). Darkness beyond the light: Attentional inhibition surrounding the classic spotlight. *NeuroReport*, *13*, 773–778.

Slotnick, S. D., Schwarzbach, J., & Yantis, S. (2003). Attentional inhibition of visual processing in human striate and extrastriate cortex. *NeuroImage*, *19*, 1602–1611.

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Q. J. Exp. Psychol.*, *40A*, 201–237.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.*, *12*, 97–136.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *J. Exp. Psychol. Hum. Percept. Perform.*, *16*, 459–478.

Treue, S., & Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*, 575–579.

Tsal, Y., & Lamy, D. (2000). Attending to an object's color entails attending to its location: Support for location-special views of visual attention. *Percept. Psychophys.*, *62*, 960–968.

Tsotsos, J. K. (2005). The selective tuning model for visual attention. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.) *Neurobiology of attention* (pp. 562–569). San Diego: Elsevier.

Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., & Pinilla, T. (1998). Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *J. Cogn. Neurosci.*, *10*, 137–151.

Vecera, S. P. (1997). Grouped arrays versus object-based representations: Reply to Kramer et al. (1997). *J. Exp. Psychol. Gen.*, *126*, 14–18.

Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *J. Exp. Psychol. Gen.*, *123*, 146–160.

Vogel, E. K., & Luck, S. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*, 190–203.

Wannig, A., Rodriguez, V., & Freiwald, W. A. (2007). Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron*, *54*, 639–651.

Weber, T. A., Kramer, A. F., & Miller, G. A. (1997). Selective processing of superimposed objects: An electrophysiological analysis of object-based attentional selection. *Biol. Psychol.*, *45*, 159–182.

Wijers, A. A., Mulder, G., Okita, T., Mulder, L. J. M., & Scheffers, M. K. (1989). Attention to color: An analysis of selection, controlled search, and motor activation, using event-related potentials. *Psychophysiology*, *26*, 89–109.

Woldorff, M. G., Fox, P. T., Matzke, M., Lancaster, J. L., Veeraswamy, S., Zamarripa, F., et al. (1997). Retinotopic organization of early visual spatial attention effects as revealed by PET and ERSs. *Hum. Brain Mapp.*, *5*, 280–286.

Wolfe, J. M., & Bennet, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vis. Res.*, *37*, 25–43.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.*, *15*, 419–433.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.*, *5*, 495–501.

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, *400*, 867–869.

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *J. Exp. Psychol. Hum Percept. Perform.*, *29*, 121–138.

16 Integration of Conflict Detection and Attentional Control Mechanisms: Combined ERP and fMRI Studies

GEORGE R. MANGUN, CLIFFORD D. SARON, AND BONG J. WALSH

ABSTRACT Attention involves powerful top-down mechanisms for the control of information processing in the brain, including specialized systems in the frontal and parietal cortex. These networks for attentional control are sensitive to momentary goals, enabling flexibility in behavior under changing conditions. One such influence is that which arises when competing inputs or responses are in conflict. A prominent system for conflict detection, cognitive control, and behavioral adjustment involves the anterior cingulate cortex and dorsal lateral prefrontal cortex. Here we describe how these attentional and conflict-resolution systems interact, and provide a synthesis and model of how these interactions support information processing. We show that conflict detected in the anterior cingulate system influences the activity of the frontal-parietal attention network to modulate attentional selection. As a result, when conditions lead to uncertainty that could be detrimental to successful performance, the brain combines information to strategically alter performance on a moment-to-moment basis.

Visual selective attention is a powerful cognitive ability that aids in the perception of the world around us (see Treisman, chapter 12 in this volume). Directing spatial attention to a location in the visual field facilitates processing of stimuli appearing at the attended location: reaction times (RT) are faster and discrimination accuracy is enhanced for events at attended versus unattended locations (e.g., Luck et al., 1994). In line with this observed RT pattern, neural responses to attended and ignored stimuli are modulated (see Maunsell, chapter 19) to provide a competitive advantage for attention events (see Kastner, McMains, and Beck, chapter 13).

Attentional control

Models of attention have distinguished between top-down and bottom-up influences on the focus of attention. In one prominent framework, the "sources" of attentional control and the resultant influence at a "site" of action, such as within the perceptual system, outline the push-pull between top-down and bottom-up information (e.g., Posner & Petersen, 1990; Serences et al., 2005). The concept that attention involves the interactions of neural systems that generate *attentional control signals* with other systems that are influenced by those signals remains at the core of most current models of voluntary (goal-directed) attention (e.g., Bundesen, Habekost, & Kyllingsbaek, 2005).

Research in animals, patients with neurological dysfunction, and healthy human subjects using electromagnetic recording, neuronal stimulation, deactivation, neuroimaging, and transcranial magnetic stimulation suggests that voluntary control of visual attention involves a complex network of widely distributed areas, including superior frontal cortex, posterior parietal cortex, posterior-superior temporal cortex, and thalamic and midbrain structures (e.g., Bisley & Goldberg, 2006; Bushnell, Goldberg, & Robinson, 1981; Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Thiebaut de Schotten et al., 2005; Goldberg & Bruce, 1985; Hopf & Mangun, 2000; Hopfinger, Buonocore, & Mangun, 2000; Hung, Driver, & Walsh, 2005; Knight, Grabowecky, & Scabini, 1995; McAlonan, Cavanaugh, & Wurtz, 2006; Mesulam, 1981; Miller, 2000; Rorden, Fruhmann Berger, & Karnath, 2006). In humans, the functional anatomy of this attentional control network has been identified by combining event-related fMRI methods with tasks that temporally separate preparatory attentional control from target-related activity. For example, using spatial cuing paradigms, it has been possible to demonstrate activity in a frontal-parietal attention system related to top-down attentional control (e.g., Corbetta et al., 2000; Hopfinger et al., 2000; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999) and to distinguish this activity from activity in visual cortex and the motor system (see Corbetta, Sylvester, & Shulman, chapter 14).

GEORGE R. MANGUN, CLIFFORD D. SARON, and BONG J. WALSH Center for Mind and Brain, University of California, Davis, California

The frontal-parietal attention network has been observed to be active in attentional control for *covert* spatial attention (Corbetta et al., 2000, 2005; Corbetta & Shulman, 2002; Giesbrecht, Woldorff, Song, & Mangun, 2003; Hopfinger et al., 2000; Kastner et al., 1999; Lepsien, Griffin, Devlin, & Nobre, 2005; Sapir, d'Avossa, McAvoy, Shulman, & Corbetta, 2005; Slagter et al., 2007), *overt* spatial attention (Astafiev et al., 2003), visual search (Shulman et al., 2003), nonspatial feature-based attention (e.g., Giesbrecht et al., 2003; Shulman, d'Avossa, Tansy, & Corbetta, 2002), global and local levels of hierarchical stimuli (Weissman, Mangun, & Woldorff, 2002; Weissman, Woldorff, Hazlett, & Mangun, 2002; Weissman, Giesbrecht, Song, Mangun, & Woldorff, 2003), and orienting attention to internal mental representations (Lepsien & Nobre, 2006). Highly similar regions of superior frontal and parietal cortex appear to be activated across studies. However, it has been proposed that attentional control networks might also involve specialization, with a subset of the attention network being highly specialized for the control of spatial attention in comparison to nonspatial attention (Giesbrecht et al., 2003; Slagter et al., 2007).

Conflict and cognitive control

Voluntary attentional control is part of the larger domain of *executive control* that includes a variety of high-level functions (e.g., Baddeley, 1996). A key component of executive control involves monitoring task performance and adjusting strategies appropriately (Gratton, Coles, & Donchin, 1992). A prominent model in this regard is the *conflict-control* model by Carter and Cohen and their colleagues (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Carter et al., 1998). The conflict-control model focuses heavily on the role of executive control systems in overcoming prepotent responses that generate response conflict during task performance. Such is the case in the Stroop task, for example, where color names (e.g., RED) written in different ink colors (e.g., green ink) are presented and the subject is required to name the ink color (Stroop, 1935). The incompatible information in the color name leads to interference and response conflict. As a result, when an incompatible trial is presented (trial N), on the subsequent trial (trial $N+1$) an increase in executive control leads to a decrease in the errors produced by incompatible information and an overall slowing in responding. This pattern of results is a cornerstone of contemporary research on executive control when response conflict is present in a task.

The conflict-control model proposes that the anterior cingulate cortex (ACC) monitors the amount of response conflict on a given trial, temporarily increasing the strength of executive control when conflict is detected (Carter et al., 1998). The conflict-control model further posits an interaction between the ACC and the dorsolateral prefrontal cortex (DLPFC), an area widely implicated in executive control processes (D'Esposito et al., 1995; Miller & Cohen, 2001). Both a dissociation and a possible functional interaction between these interconnected brain regions have been demonstrated. MacDonald, Cohen, Stenger, and Carter (2000) showed that the activity in DLPFC but not ACC increased when a "color-naming" cue (as compared with a "word-reading" cue) was presented during a delayed Stroop task, whereas ACC but not DLPFC activity increased in response to the presentation of the incongruent stimulus (compared to a congruent stimulus). The idea supported by such findings is that the conflict-generated activity in ACC could signal the need for increased cognitive control that would be exerted by DLPFC.

A study from Carter and colleagues provided more direct evidence for this neural model by demonstrating "next-trial effects" in DLPFC: High-conflict trials resulted in increased ACC activity on the current trial (trial N), leading to increased DLPFC activity for the subsequent trial (trial $N+1$), and this corresponded with improved behavioral performance (fewer errors) on those subsequent trials (Kerns et al., 2004). Evidence for a causal relationship between DLPFC activity and attentional selection in perceptual processing comes from the report by Egner and Hirsch (2005), who used a modified Strooplike task containing the names and faces of famous persons (actors and politicians). They reported that when subjects were categorizing faces, the activity in the fusiform face area (FFA) was enhanced for trials following a high-conflict trial, and that there was also increased functional connectivity between the DLPFC and the FFA under such conditions. These results fit nicely with the conflict-control model, which proposes that ACC activity is related to the detection of conflict, which then signals the need for increased cognitive control that is implemented by the DLPFC. Further, the pattern of Egner and Hirsch suggests that DLPFC activity is correlated with the results of selective attention in perceptual systems.

Conflict and attentional control systems

Interrelationships between the ACC-DLPFC conflict-control system and the frontal-parietal attentional control system might explain how conflict detection ultimately leads to selective stimulus processing. For example, Casey and colleagues (2000) provided evidence that incompatible trials may lead to increased attentional control in tasks where selective attention to targets relative to distracters might minimize interference from irrelevant or incompatible information. They were able to demonstrate that superior frontal and superior parietal cortex (attentional control network) was activated when successive incompatible trials were presented. The idea suggested by this research is that selective attention was engaged to focus attention on the relevant input and suppress

distracter information, much in the way suggested by Egner and Hirsch (2005). Using event-related fMRI, Weissman, Mangun, and Woldorff (2002) demonstrated similar activations in the frontal-parietal attention network to attention-directing cues and subsequent targets containing incompatible local and global information. This type of information can be interpreted as evidence that during the incompatible trials, the voluntary attention system was engaged to focus attention on the information at the relevant level (global or local) and suppress the irrelevant, distracting information. Together these studies suggest a close association between the conflict-control system (ACC-DLPFC) and attentional control networks (frontal-parietal network). However, there is less direct evidence that increased conflict leads to modulations of activity in the frontal-parietal attentional control network. To demonstrate such a relationship, it would be necessary to show that increased conflict leading to activity in the ACC on a trial (trial *N*) would lead to increased activity in the frontal-parietal attention network on the subsequent trial (trial $N + 1$). Further, such effects should be strongly correlated with attentional selectivity demonstrated behaviorally and in selective stimulus processing in visual cortex.

Integrating brain networks for conflict processing and attention

Using functional magnetic resonance imaging (fMRI), it was investigated whether the ACC conflict-detection system interacts with the frontal-parietal attentional control system when spatial cuing leads to conflicts in attentional orienting (Walsh, 2008). The working hypothesis was that conflict in one trial should result in adjustments in frontal-parietal top-down attentional control to reduce conflict in a subsequent trial.

The paradigm was a hybrid of those used in conflict-controls studies and those typical of spatial attention studies (figure 16.1). Cues consisting of two vertical lines directed the subjects' covert spatial attention (100% instructive) to either the right or the left visual in order to discriminate the features of an upcoming target. The longer of the two lines in the cue indicated that spatial attention should be directed to the hemifield on the side of the longer line. The amount of conflict generated by the cues was systematically manipulated by varying the difference in length of two lines (see figure 16.1*B*), in line with evidence that requiring near-threshold-level perceptual judgments can result in the generation of conflict (Szmalec et al., 2008). As a result, low, medium, and high conflict-generating cues were created. Bilateral targets (200 ms duration) followed the cues after a delay period (1500 ms), and these targets were then masked at offset by pattern masks (300 ms duration). Subjects were required to respond by pressing one of two buttons with their right hand

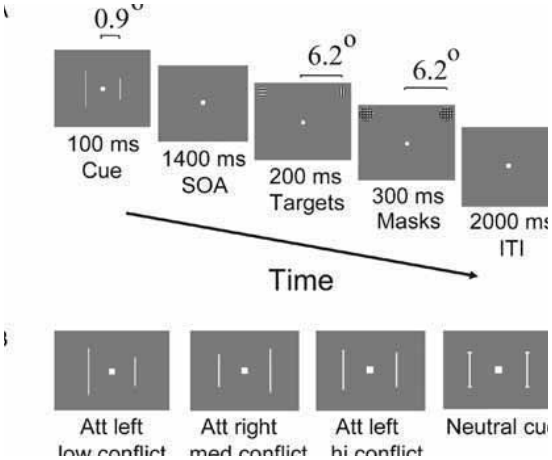


FIGURE 16.1 (*A*) Stimulus sequence. Subjects received a cue that directed attention to left or right, or to neither side (neutral cues). Following a cue-to-target stimulus onset asynchrony (SOA), targets were presented. Targets were circular Gabor gratings (1.5° diameter, located bilaterally 5.4° from fixation in the upper left and right visual fields), each of which could be horizontal or vertical in orientation. Subjects had to discriminate the grating orientation at the attended location. (*B*) Sample cue types. The length of each line ranged from 1.1° to 1.7°, depending on the type of cue. Neutral cues (to which subjects did not shift attention) were similar to attend cues, except the vertical lines were of equal length (1.1°) and had short (0.2°) horizontal lines on the top and bottom of each vertical line to distinguish them from attend cues. Subjects were instructed to shift attention to the hemifield indicated by the cue (shift left if left line is longer, shift right if right line is longer). They were told to respond with a button press indicating if the gratings of the subsequent target were vertical or horizontal.

whether the target in the attended field was horizontal or vertical (attended and unattended targets could both be the same orientation or each be different orientations).

As has been observed in numerous prior studies that have used signal-processing methods to decompose cue from target activity (e.g., Ollinger, Shulman, & Corbetta, 2001; Woldorff et al., 2004) in studies of spatial attention (e.g., Corbetta et al., 2000; Hopfinger et al., 2000; see Corbetta et al., chapter 14, this volume, for a review), the frontal-parietal attentional control system was activated in response to the attention-directing cues (collapsed across level of conflict) (figure 16.2*A*), but not to targets. In addition, in visual cortical regions, in response to cues, there was a significant activation of the visual cortex contralateral to the direction attention was cued (left versus right) (figure 16.2*B*). Targets also activated visual cortex, and these sensory-evoked activations were modulated by the direction of covert spatial attention such that responses to the bilateral targets were larger on the hemisphere contralateral to the attended hemifield (not shown in figures). Therefore, independent of cue conflict, the frontal-parietal attentional control network was activated when subjects acted on the cue instructions, and this activation led to changes in visual cortex.

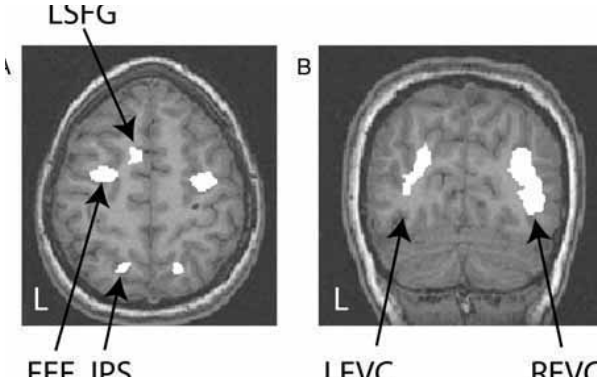


FIGURE 16.2 (*A*) Frontal-parietal attentional control network defined by the attend-cue versus neutral-cue contrast, collapsed over cue-left and cue-right and levels of conflict. Attentional control activity was observed in bilateral frontal eye fields (FEF), bilateral posterior parietal cortex in and around the intraparietal sulcus (IPS), and the superior frontal gyrus in the left hemisphere (LSFG). (*B*) Increased baseline activity in visual cortex time-locked to the attention cues. The contrast is for cue-left versus cue-right, where activity is greatest in the visual extrastriate visual cortex contralateral to the cued hemifield. That is, cues directing attention to the right hemifield resulted in increased activity in left extrastriate visual cortex (LEVC), whereas cues to the left resulted in increased activity in the right extrastriate visual cortex (REVC).

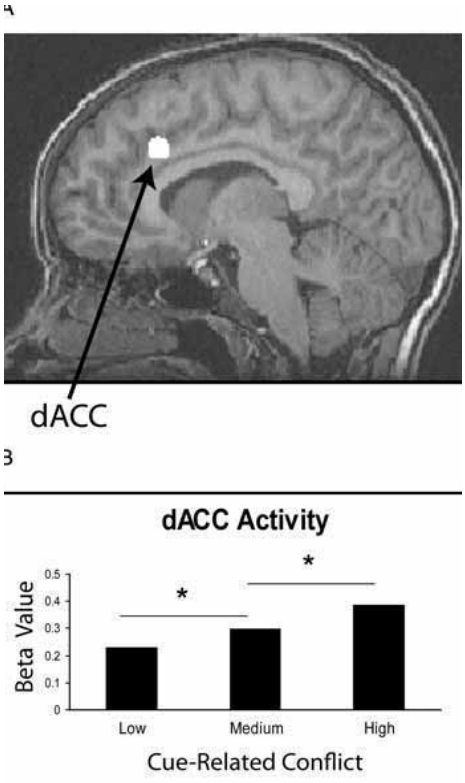


FIGURE 16.3 (*A*) Activity in dorsal anterior cingulate cortex (dACC) related to cue-induced conflict (contrast between high- and low-conflict cue trials—see figure 16.1*B*). (*B*) Within the dACC region of interest (ROI) shown in *A*, there was a parametric increase in fMRI BOLD signal (plotted as beta values) as function of increasing cue-related conflict.

The goal in this work was to investigate whether difference in conflict in attentional orienting (a function of cue discriminability) led to systematic changes in the frontal-parietal control regions. This line of argument, however, is dependent on the paradigm and cue discriminability manipulation resulting in conflict that activates the ACC and triggers cognitive control. To establish this relationship, it is important to demonstrate that the cue manipulation resulted in standard effects on behavior and brain activity (e.g., Kerns et al., 2004). Analysis of the reaction times (RT) and accuracy to detect targets at the cued locations were broken down as a function of whether the trial was preceded by a high versus low conflict trial (defined by whether the cue was high versus low conflict in the preceding trial). Reaction times were faster ($p < .001$) and accuracy was higher ($p < .005$) for detecting targets when preceded by high-conflict trials versus low-conflict trials. Thus the expected pattern of improved speed and accuracy of behavioral responses in trials following high versus low conflict trials was observed in this spatial cuing paradigm. This behavioral signature suggests that conflict resulted in increased attentional control that was manifest as improvements in behavior (e.g., Kerns et al., 2004).

Next, whole-brain voxel-wise analyses of the fMRI BOLD responses were conducted to determine whether the observed behavioral adjustments were mediated by the well-known mechanisms involving the ACC. Activity associated with high-conflict cues compared to activity associated with low-conflict attend cues was found to include the dorsal anterior cingulate cortex (dACC), as well as small regions of DLPFC, right lateral parietal cortex, and left anterior insula (figure 16.3). Within the dACC region identified by the contrast, the amplitudes of BOLD signals were observed to be parametrically related to degree of conflict engendered by the cues (i.e., as a function of cue discriminability). These findings were in line with numerous prior studies of the role of the ACC in conflict detection and indicate that the cue manipulation was effective in activating the ACC (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; MacDonald et al., 2000; Liston, Matalon, Hare, Davidson, & Casey, 2006).

Finally, electrophysiology measures have previously been related to conflict detection and should therefore be elicited in the present hybrid design. Studies using ERP have identified an anterior midline negativity (referred to as the N2 component) that is generated in the ACC and increases in amplitude to stimuli that are more likely to induce conflict (van Veen & Carter, 2002a, 2002b; Donkers & van Boxtel, 2004; Szmalec et al., 2008). In a subset of the same subjects tested in the fMRI study, ERPs were recorded from 128 channels as they performed the task shown in figure 16.1 (Walsh, 2008). As expected, the amplitude of the N2 component of the evoked potential at frontocentral electrode

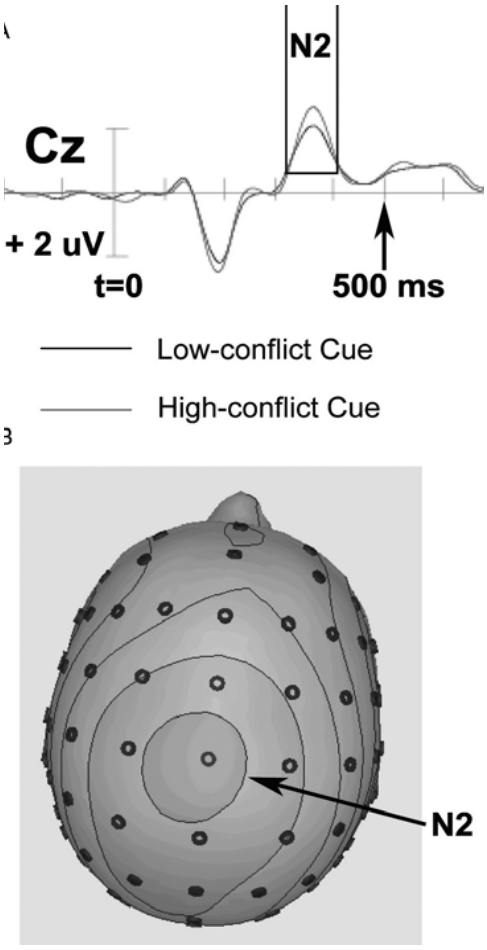


FIGURE 16.4 (*A*) Grand average ERPs to cues as a function of conflict, recorded from midline central scalp electrode site Cz. Negative voltage is plotted upward, and cue onset is the upright bar at time zero. High-conflict cue ERPs are plotted in red. The box indicates the N2 component, peaking at approximately 360 ms from cue onset. The N2 is significantly greater in amplitude for high-conflict cues. (*B*) Scalp topography at the peak of the N2 response showing the midline central scalp maximum of the response (blue colors). The nose is at the top of the figure of the scalp, and left is on the left of the image. The small red circles are the locations of the electrodes. (See color plate 18.)

sites reflected the degree of conflict engendered by the cues. The N2 was larger in response to high-conflict cues (high > low conflict cue: FCz/Cz, $p < .02$) (figure 16.4).

CONFLICT INDUCES MODULATIONS OF THE FRONTAL-PARIETAL CONTROL NETWORK The behavioral, fMRI, and ERP data provided evidence that changes in cue discriminability resulted in conflict that activated the dACC, affecting performance on the next trial. If activity in the dACC triggers increased attentional control, we might expect to observe those changes on the next trial in the frontal-parietal attention network. To investigate this question, the activations identified by contrasts of attention cues to neutral cues were used as regions of interest (ROIs). The hemodynamic responses in these regions of interest could then be investigated as a function of attention or conflict in the current trial (trial *N*) and the next trial (trial *N* + 1). These results are shown in figure 16.5 for the dACC, the FEF, and the intraparietal cortex. Two main findings can be highlighted.

First, as described earlier, contrasts related to attention revealed activity in the frontal-parietal network but not the dACC, whereas contrasts related to conflict affected the dACC but not the frontal-parietal network (figure 16.5). This pattern draws the distinction between the dACC that is sensitive to cue conflict (e.g., Kerns et al., 2004) and the frontal-parietal network that is sensitive to attentional control (e.g., Corbetta et al., 2000; Hopfinger et al., 2000).

The second finding is the pattern of attention and conflict in the dACC and the frontal-parietal network for trial *N* versus trial *N* + 1. This can be observed in figure 16.5 by observing the time courses of the hemodynamic responses. The key finding is that although cue conflict does not result in fMRI BOLD signal changes in the frontal-parietal network for trial *N*, in response to trial *N* + 1 (following a high-conflict cue) this network shows a robust response. This pattern shows that high (versus low) cue conflict on one trial leads to increased activity in the frontal-parietal network on the next trial. This pattern is consistent with that observed in behavior where target performance was improved on trials that followed a high-cue-conflict trial (described earlier).

Conclusions

The results of these fMRI and ERP studies support the idea that conflict-monitoring networks and attentional-control networks interact to modulate spatial attention to improve performance. High attentional cue conflict in one trial (trial *N*) generated high ACC activity in that trial, and then led to increased frontal-parietal activity in the next trial (trial *N* + 1). Such a pattern of results establishes a long-hypothesized relationship between conflict and control mechanisms and selective attention (e.g., Casey et al., 2000). Stated more generally, a performance-monitoring system that detects conflict related to orienting attention appears to signal attentional control systems to exert greater control over the focus of attention (Botvinick et al., 2001; Kerns et al., 2004; Egner & Hirsch, 2005; Liston et al., 2006). The interactions of these systems are diagrammed in figure 16.6, which shows how conflict signals the need for attentional control that leads to changes in behavior.

The neural substrates for such interactions between conflict and attention systems have been reported. The ACC has been shown to project directly to both frontal eye fields (FEF) (Huerta, Krubitzer, & Kass, 1987; Stanton, Bruce, & Goldberg, 1993) and posterior parietal cortex

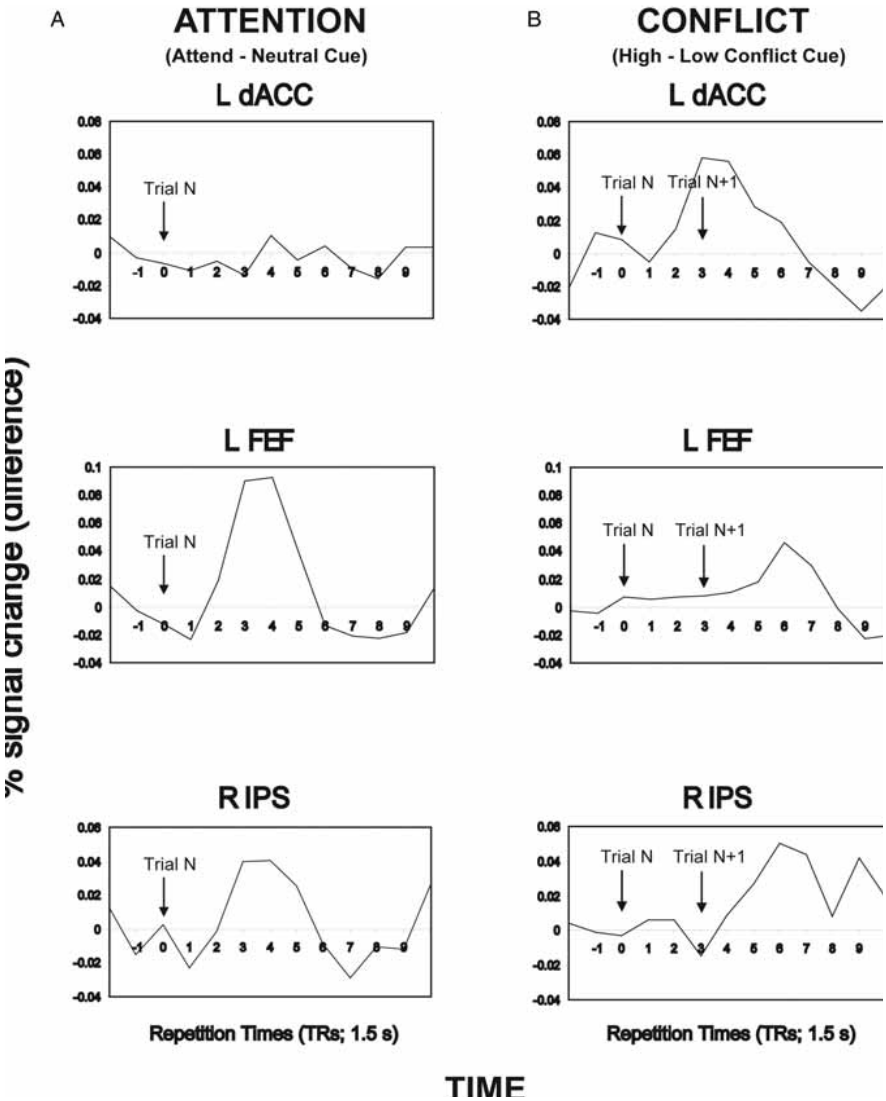


FIGURE 16.5 Hemodynamic responses from dorsal anterior cingulate cortex (dACC), frontal eye fields (FEF), and the intraparietal sulcus (IPS) in parietal cortex. The plots represent differences in percent signal change computed by subtracting two conditions. (*A*) Time courses for attend cues minus neutral cues (the effects of attentional control) show activity in the FEF and IPS but not in the dACC. (*B*) Time courses for high-conflict cues minus low-conflict cues (the effects of conflict) show that for the current trial (trial *N*) the dACC was active while the FEF and IPS were not. In contrast, however, the reverse pattern was true for trial $N + 1$ where a hemodynamic response is observed that is delayed, being time-locked to trial $N + 1$ in the FEF and IPS (but not dACC).

(Pandya, Van Hoesen, & Mesulam, 1981; Selemon & Goldman-Rakic, 1988), critical elements of the frontal-parietal attentional control network. Given this connectivity, and the results obtained in the experiments described in this chapter, one might speculate that the ACC projects directly to structures in the frontal-parietal attention network critical to modulating attentional control.

Prior research in cognitive control has established a relationship between the ACC and the dorsolateral prefrontal cortex (DLPFC) where conflict resulting in ACC activation triggered DLPFC activity that was related to changes in performance (e.g., Kerns et al., 2004). In the studies reported in this chapter, no such DLPFC involvement was measurable, raising the possibility that for attentional conflict, interactions between ACC and the frontal-parietal attention network were not mediated by the DLPFC; this must remain a hypothesis for future investigation.

The present formulation can be considered in relation to other models of cortical control systems. One such model is the recent proposal regarding the interaction of cortical networks in default-mode processing and cognitive control. Dosenbach, Fair, Cohen, Schlagger, and Petersen (2008) proposed a model in which a frontal-parietal system is involved in top-down control processes that meet moment-to-moment demands, while a cingulate-opercular system (involving ACC, anterior prefrontal cortex, and inferior

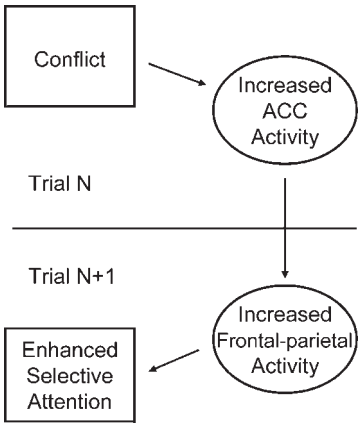


FIGURE 16.6 Diagram of the interactions of the conflict and attention systems.

lateral prefrontal cortex, as well as subcortical structures) is involved in maintaining a stable set over time during task performance. They noted that "it seems likely that additional controllers might exist, operating at other temporal and/or spatial scales" (p. 103). The findings reviewed in the present chapter emphasize this general notion of interacting cortical networks that act to achieve different specific computational needs during performance. By showing that systems involved in conflict monitoring and error processing interact in dynamic ways with voluntary attentional control networks to modulate selective attention, the present work establishes the interactions of two major cognitive control systems for evaluating, controlling, and improving perception and performance.

ACKNOWLEDGMENTS We are deeply grateful to Cameron S. Carter, Michael H. Buonocore, Barry Giesbrecht, Sean P. Fannon, and Dorothee Heipertz for their collaboration, advice, and assistance. Supported by NIMH R01 MH55714 to GRM and NEI Traineeship T32 EY015387 to BJW.

REFERENCES

ASTAFIEV, S. V., SHULMAN, G. L., STANLEY, C. M., SNYDER, A. Z., VAN ESSEN, D. C., & CORBETTA, M. (2003). Functional organization of human intraparietal and frontal cortex for attending, looking, and pointing. *J. Neurosci.*, *23*(11), 4689–4699.

BADDELEY, A. (1996). Exploring the central executive. *Q. J. Exp. Psychol.* [*A*], *49A*(1), 5–28.

BISLEY, J. W., & GOLDBERG, M. E. (2006). Neural correlates of attention and distractibility in the lateral intraparietal area. *J. Neurophysiol.*, *95*, 1696–1717.

BOTVINICK, M. M., BRAVER, T. S., BARCH, D. M., CARTER, C. S., & COHEN, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.*, *108*(3), 624–652.

BOTVINICK, M., NYSTROM, L. E., FISSELL, K., CARTER, C. S., & COHEN, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*(6758), 179–181.

BUNDESEN, C., HABEKOST, T., & KYLLINGSBAEK, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychol. Rev.*, *112*(2), 291–328.

BUSHNELL, M. C., GOLDBERG, M. E., & ROBINSON, D. L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *J. Neurophysiol.*, *46*(4), 755–772.

CARTER, C. S., BRAVER, T. S., BARCH, D. M., BOTVINICK, M. M., NOLL, D., & COHEN, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*(5364), 747–749.

CASEY, B. J., THOMAS, K. M., WELSH, T. F., BADGAIYAN, R. D., ECCARD, C. H., JENNINGS, J. R., et al. (2000). Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA*, *97*(15), 8728–8733.

CORBETTA, M., KINCADE, J. M., OLLINGER, J. M., MCAVOY, M. P., & SHULMAN, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.*, *3*, 292–297.

CORBETTA, M., & SHULMAN, G. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.*, *3*(3), 201–215.

CORBETTA, M., TANSY, A. P., STANLEY, C. M., ASTAFIEV, S. V., SNYDER, A. Z., & SHULMAN, G. L. (2005). A functional MRI study of preparatory signals for spatial location and objects. *Neuropsychologia*, *43*(14), 2041–2056. Epub 2005 Apr 26.

D'ESPOSITO, M., DETRE, J. A., ALSOP, D. C., SHIN, R. K., ATLAS, S., & GROSSMAN, M. (1995). The neural basis of the central executive system of working memory. *Nature*, *378*, 279–281.

DONKERS, F., & VAN BOXTEL, G. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain Cogn.*, *56*(2), 165–176.

DOSENBACH, N. U., FAIR, D. A., COHEN, A. L., SCHLAGGAR, B. L., & PETERSEN, S. E. (2008). A dual-networks architecture of top-down control. *Trends Cogn Sci.*, *12*(3):99–105.

EGNER, T., & HIRSCH, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat. Neurosci.*, *8*(12), 1784–1790.

GIESBRECHT, B., WOLDORFF, M. G., SONG, A. W., & MANGUN, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *NeuroImage*, *19*(3), 496–512.

GOLDBERG, M. E., & BRUCE, C. J. (1985). Cerebral cortical activity associated with the orientation of visual attention in the rhesus monkey. *Vision Res.*, *25*(3), 471–481.

GRATTON, G., COLES, M. G., & DONCHIN, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *J. Exp. Psychol. Gen.*, *121*(4):480–506.

HOPF, J. M., & MANGUN, G. R. (2000). Shifting visual attention in space: An electrophysiological analysis using high spatial resolution mapping. *Clin. Neurophysiol.*, *111*, 1241–1257.

HOPFINGER, J. B., BUONOCORE, M. H., & MANGUN, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat. Neurosci.*, *3*, 284–291.

HUERTA, M. F., KRUBITZER, L. A., & KASS, J. H. (1987). Frontal eye field as defined by intracortical microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys. II. Cortical connections. *J. Comp. Neurol.*, *265*(3), 332–361.

HUNG, J., DRIVER, J., & WALSH, V. (2005). Visual selection and posterior parietal cortex: Effects of repetitive transcranial magnetic stimulation on partial report analyzed by Bundesen's theory of visual attention. *J. Neurosci.*, *25*(42), 9602–9612.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human cerebral cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*, 751–761.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026.

Knight, R. T., Grabowecky, M. F., & Scabini, D. (1995). Role of human prefrontal cortex in attention control. *Adv. Neurol.*, *66*, 21–34.

Lepsien, J., Griffin, I. C., Devlin, J. T., & Nobre, A. C. (2005). Directing spatial attention in mental representations: Interactions between attentional orienting and working-memory load. *NeuroImage*, *26*(3), 733–743.

Lepsien, J., & Nobre, A. C. (2006). Cognitive control of attention in the human brain: Insights from orienting attention to mental representations. *Brain Res., 1105*(1), 20–31.

Liston, C., Matalon, S., Hare, T. A., Davidson, M. C., & Casey, B. J. (2006). Anterior cingulate and posterior parietal cortices are sensitive to dissociable forms of conflict in a task-switching paradigm. *Neuron*, *50*(4), 643–653.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. *J. Exp. Psychol. Hum. Percept. Perform.*, *20*, 887–904.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*(5472), 1835–1838.

McAlonan, K., Cavanaugh, J., & Wurtz, R. H. (2006). Attentional modulation of thalamic reticular neurons. *J. Neurosci.*, *26*(16), 4444–4450.

Mesulam, M.-M. (1981). A cortical network for directed attention and unilateral neglect. *Ann. Neurol.*, *10*, 309–325.

Miller, E. K. (2000). The neural basis of top-down control of visual attention in the prefrontal cortex. In J. Driver & S. Monsell (Eds.), *Attention and Performance XVIII: The control over cognitive processes* (pp. 511–534). Cambridge, MA: MIT Press.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.*, *24*, 167–202.

Ollinger, J. M., Shulman, G. L., & Corbetta, M. (2001). Separating processes within a trial in event-related functional MRI. *NeuroImage*, 13, 210–217.

Pandya, D. N., Van Hoesen, G. W., & Mesulam, M. M. (1981). Efferent connections of the cingulate gyrus in the rhesus monkey. *Exp. Brain Res., 42*(3–4), 319–330.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.*, *13*, 25–42.

Rorden, C., Fruhmann Berger, M., & Karnath, H. O. (2006). Disturbed line bisection is associated with posterior brain lesions. *Brain Res. Cogn. Brain Res., 1080*(1), 17–25.

Sapir, A., d'Avossa, G., McAvoy, M., Shulman, G. L., & Corbetta, M. (2005). Brain signals for spatial attention predict performance in a motion discrimination task. *Proc. Natl. Acad. Sci. USA*, *102*(49), 17810–17815.

Selemon, L. D., & Goldman-Rakic, P. S. (1988). Common cortical and subcortical targets of the dorsolateral prefrontal and posterior partietal cortices in the rhesus monkey: Evidence for a distributed neural network subserving spatially guided behavior. *J. Neurosci., 8*(11), 4049–4068.

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychol. Sci., 16*(2), 114–122.

Shulman, G. L., d'Avossa, G., Tansy, A. P., & Corbetta, M. (2002). Two attentional processes in the parietal lobe. *Cereb. Cortex*, *12*(11), 1124–1131.

Shulman, G. L., McAvoy, M. P., Cowan, M. C., Astafiev, S. V., Tansy, A. P., d'Avossa, G., et al. (2003). Quantitative analysis of attention and detection signals during visual search. *J. Neurophysiol.*, *90*(5), 3384–3397.

Slagter, H. A., Giesbrecht, B., Kok, A., Weissman, D. H., Kenemans, J. L., Woldorff, M. G., et al. (2007). fMRI evidence for both generalized and specialized components of attentional control. *Brain Res.*, *1177*, 90–102.

Stanton, G. B., Bruce, C. J., & Goldberg, M. E. (1993). Topography of projections to the frontal lobe from the macaque frontal eye fields. *J. Comp. Neurol., 330*(2), 286–301.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.*, *18*, 643–662.

Szmalec, A., Verbruggen, F., Vandierendonck, A., De Baene, W., Verguts, T., & Notebaert, W. (2008). Stimulus ambiguity elicits response conflict. *Neurosci. Lett.*, 435(2), 158–162.

Thiebaut de Schotten, M., Urbanski, M., Duffau, H., Volle, E., Levy, R., Dubois, B., & Bartolomeo, P. (2005). Direct evidence for a parietal-frontal pathway subserving spatial awareness in humans. *Science*, *309*(5744), 2226–2228. Erratum in: *Science*, *317*(5838), 597.

van Veen, V., & Carter, C. S. (2002a). The timing of action-monitoring processes in the anterior cingulate cortex. *J. Cogn. Neurosci.*, *14*(4), 593–602.

van Veen, V., & Carter, C. S. (2002b). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiol. Behav.*, *77*(4–5), 477–482. Review.

Walsh, B. J. (2008). *Characterization of trial-to-trial adjustments in selective attention as a consequence of conflict detection*. Unpublished doctoral dissertation, University of California, Davis.

Weissman, D. H., Giesbrecht, B. G., Song, A. W., Mangun, G. R., & Woldorff, M. (2003). Conflict monitoring in the human anterior cingulate cortex during selective attention to global and local object features. *NeuroImage*, *19*(4), 1361–1368.

Weissman, D. H., Mangun, G. R., & Woldorff, M. G. (2002). A role for top-down attentional orienting during interference between global and local aspects of hierarchical stimuli. *NeuroImage*, *17*, 1266–1276.

Weissman, D. H., Woldorff, M. G., Hazlett, C. J., & Mangun, G. R. (2002). Effects of practice on executive control investigated with fMRI. *Cogn. Brain Res.*, *15*, 47–60.

Woldorff, M. G., Hazlett, C. J., Fichtenholtz, H. M., Weissman, D. H., Anders, A. M., & Song, A. W. (2004). Functional parcellation of attentional control regions of the brain. *J. Cogn. Neurosci.*, *16*, 149–165.

17 A Right Perisylvian Neural Network for Human Spatial Orienting

HANS-OTTO KARNATH

ABSTRACT Homologous neural networks seem to exist in the human left and right hemispheres tightly linking cortical regions straddling the sylvian fissure. White matter fiber bundles connect the inferior parietal lobule with the ventrolateral frontal cortex, ventrolateral frontal cortex with superior/middle temporal cortex, and superior/middle temporal cortex with the inferior parietal lobule. It is argued that these perisylvian networks serve different cognitive functions, a representation for language and praxis in the left hemisphere and a representation for processes involved in spatial orienting in the right. The tight perisylvian anatomical connectivity between superior/middle temporal, inferior parietal, and ventrolateral frontal cortices might explain why lesions at these distant cortical sites around the sylvian fissure in the human right hemisphere can lead to the same disturbance of orienting behavior, namely, to spatial neglect.

In recent years it has been shown that functional and structural lateralization of the brain is more widespread among vertebrates than previously believed. Nevertheless, it is still appropriate that many motor, sensory, and visual, but also cognitive, functions show bihemispheric representations in the human and nonhuman primate. Only a few (so-called higher) cognitive functions have obvious asymmetrical representations. Among them are language, praxis, and spatial orienting. While an elaborate representation for language and praxis has evolved in the human left hemisphere, a neural system involved in spatial orienting is dominantly represented in the right hemisphere. Consequently, locally corresponding damage to one of the two hemispheres leads to different symptoms. While the dominant disorders in neurological patients with left hemisphere involvement are aphasia and apraxia, patients with right hemisphere damage typically show spatial neglect. This term describes a spontaneous deviation of the eyes and the head toward the ipsilesional, right side (Fruhmann-Berger & Karnath, 2005; Fruhmann-Berger, Pross, Ilg, & Karnath, 2006). Patients with such a disorder disregard objects located on the contralesional, left side. When searching for targets, copying, or reading, for example, they concentrate their exploratory movements predominantly on the right side of space (Heilman, Watson, Valenstein, & Damasio, 1983; Behrmann, Watt, Black, & Barton, 1997; Karnath, Niemeier, & Dichgans, 1998). The question thus arises whether the development of these different functions in the human left and right hemispheres corresponds with different anatomical representations. Or is it possible that homologous neural structures serve as correlates for language and praxis in the left and for spatial orientation in the right hemisphere?

Three major cortical areas have been described as neural correlates of spatial neglect in the human right hemisphere. A first study by Heilman and coworkers (1983) revealed the right inferior parietal lobule (IPL) and the temporoparietal junction (TPJ). Subsequent studies reported comparable observations (e.g., Vallar & Perani, 1986; Mort et al., 2003). Lesions located in the right ventrolateral frontal cortex were also observed to correlate with spatial neglect (Vallar & Perani, 1986; Husain & Kennard, 1996; Committeri et al., 2007). Finally, several studies have revealed the right superior temporal cortex and adjacent insula as being critically related to the disorder (Karnath, Ferber, & Himmelbach, 2001; Karnath, Fruhmann-Berger, Küker, & Rorden, 2004; Buxbaum et al., 2004; Corbetta, Kincade, Lewis, Snyder, & Sapir, 2005; Committeri et al., 2007; Sarri, Greenwood, Kalra, & Driver, 2009).

Interestingly, a similar pattern of perisylvian correlates has been observed in the human left hemisphere when stroke patients suffer from aphasia. Early analyses (e.g., Kertesz, Harlock, & Coates, 1979; Poeck, de Bleser, Graf von Keyserlingk, 1984), as well as more recent studies of cortical lesion localization in neurological patients with disorders in language comprehension and/or speech production (e.g., Kreisler et al., 2000; Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004; Borovsky, Saygin, Bates, & Dronkers, 2007), revealed involvement of the ventrolateral frontal cortex, superior and middle temporal gyri, insula, and IPL. These

HANS-OTTO KARNATH Center for Neurology, University of Tübingen, Tübingen, Germany

findings are supported by electrical mapping of the human cortex during surgery in awake patients as well as functional magnetic resonance imaging (fMRI) in healthy subjects. A recent meta-analysis of 129 fMRI studies on phonological, semantic, and syntactic processing revealed activation in distributed areas predominantly involving left middle and inferior dorsolateral frontal, superior and middle temporal, and inferior parietal cortices (Vigneau et al., 2006). These sites correspond very well with those in which intraoperative cortical stimulation evoked disturbances of language processes, such as anomia, alexia, or speech arrest (e.g., Boatman, 2004; Duffau et al., 2005; Sanai, Mirzadeh, & Berger, 2008).

These perisylvian brain areas in the human left hemisphere do not appear to represent language processes solely. Recent analyses of lesion localization in patients suffering from apraxia suggested that they are also involved in the organization of motor actions (Goldenberg & Karnath 2006; Goldenberg, Hermsdörfer, Glindeman, Rorden, & Karnath, 2007). Stroke patients with either disturbed pantomime of tool use or with disturbed imitation of finger postures typically showed damage of the left inferior frontal gyrus (IFG) and adjacent portions of the insula, while disturbed imitation of hand postures was associated with posterior lesions affecting the IPL and TPJ. This close anatomical relationship between the representation of praxis on the one hand and language on the other led to the assumption that these left perisylvian areas might represent an observation/execution matching system providing the bridge from "doing" to "communicating" (Rizzolatti & Arbib, 1998; Iacoboni & Wilson, 2006). Its development was seen as a consequence of the fact that before speech appearance the precursors of these areas in the monkey were endowed with a mechanism for recognizing actions made by others. This mechanism was seen as the neural prerequisite for the development of interindividual communication and finally of speech (Rizzolatti & Arbib, 1998).

Thus it seems as if very similar anatomical cortical areas straddling the sylvian fissure are involved in representing language and praxis in the human left hemisphere and spatial orienting in the right hemisphere. Recent anatomical studies have revealed that a dense white matter connectivity exists specifically between these perisylvian cortical areas. In the following sections, it will be argued that intimately interconnected homologous perisylvian networks have evolved in the human left and right hemispheres serving for different cognitive functions, a representation for language and praxis in the left hemisphere and a representation for spatial orienting in the right hemisphere. Further, it will be argued that for these cognitive processes the functioning of the perisylvian cortical areas is critical, not the mere disconnection of their white matter interconnections.

Dense perisylvian white matter connectivity

Beyond traditional axonal-tract-tracing and myelin-staining techniques, the development of diffusion-based imaging—for example, diffusion tensor and diffusion spectrum imaging (DTI/DSI)—has opened new opportunities for identifying long-range white matter pathways. By combining the findings from DSI and from histological tract tracing, Schmahmann and colleagues separated ten long, bidirectional association fiber bundles in the monkey brain (Schmahmann & Pandya, 2006; Schmahmann et al., 2007). Anatomical homologies have been described in the human with the aid of diffusion-based imaging in vivo (e.g., Catani, Howard, Pajevic, & Jones, 2002; Catani, Jones, & ffytche, 2005; Catani et al., 2007; Makris et al., 2005, 2007; Mori, Wakana, van Zijl, & Nagae-Poetscher, 2005; Upadhyay, Hallock, Ducros, Kim, & Ronen, 2008) and myelin staining postmortem (Bürgel et al., 2006). In the following, the focus will be on those pathways that connect the perisylvian cortical areas—that is, the superior/middle temporal, inferior parietal, and dorsolateral frontal cortices—where damage has been shown to provoke spatial neglect in the case of right brain lesions and aphasia and/or apraxia after left hemisphere involvement.

In the monkey and human, the superior longitudinal fasciculus (SFL) is the major cortical association fiber pathway linking parietal and frontal cortices. It is subdivided into different, separable components (Petrides & Pandya 1984; Makris et al., 2005; Schmahmann et al., 2007). One part—the SLF I—is situated dorsally of the perisylvian network area, connecting the superior parietal lobule with dorsal premotor areas. Two further subcomponents connect the IPL with premotor and prefrontal cortices. The SLF II links the IPL and intraparietal sulcus with the posterior and caudal prefrontal cortex, while the SLF III connects the rostral IPL with the ventral part of premotor and prefrontal cortex. A further fiber tract that is separable from these connections stems from the caudal part of the superior/middle temporal gyrus (STG/MTG), arches around the caudal end of the sylvian fissure, and extends to the lateral prefrontal cortex along with the SLF II fibers. This latter fiber tract is termed the arcuate fasciculus (AF) and has been described in humans (Burdach, 1819–26; Dejerine & Dejerine-Klumpke, 1895; Catani et al., 2002, 2005; Makris et al., 2005; Vernooij et al., 2007; Upadhyay et al., 2008) as well as in the monkey (Petrides & Pandya, 1988; Schmahmann et al., 2007). Some disagreement exists related to whether or not this fiber bundle is regarded as a fourth subdivision of the SLF (SLF IV; Makris et al., 2005; Vernooij et al., 2007), a stand-alone connection adjacent to the SLF (Schmahmann et al., 2007), or only part, namely, the long segment (discussed later), of a three-way AF structure (Catani et al., 2005).

A fiber bundle situated in close proximity to the SLF II and the AF is the superior occipitofrontal fasciculus (SOF), also termed the (superior) fronto-occipital fasciculus ([S]FOF) by some authors. The SOF forms the medial border of the corticospinal tract and separates it from the lateral ventricles. The fibers run parallel to the dorsolateral margin of the lateral ventricles below the corpus callosum. Some of its fibers intermingle with SLF II and AF fibers. Different from expectations based on its labeling, the SOF connects not only occipital but also inferior parietal with frontal lobe areas. In the monkey as well as in humans, this long association bundle bidirectionally extends from the IPL and dorsomedial parastriate occipital cortex to caudal, dorsal, and medial frontal lobe areas (Catani et al., 2002; Bürgel et al., 2006; Makris et al., 2007; Schmahmann et al., 2007).

Two further, ventrally located fiber bundles contribute to the perisylvian network focused on in this chapter, namely, the strong pathway running through the extreme capsule (EmC) and the middle longitudinal fasciculus (MdLF). Both bundles have been described in monkeys (Seltzer & Pandya, 1984; Schmahmann & Pandya, 2006; Schmahmann et al., 2007; Petrides & Pandya, 2007) as well as in humans (Makris & Pandya, 2009; Makris et al., 2009). The EmC is situated between the claustrum and the insular cortex interconnecting the inferior frontal and orbitofrontal gyri with the midportion of the superior temporal region. It further continues caudally toward the occipital cortex and toward the IPL, flanking here another fiber pathway, namely the MdLF (Makris & Pandya, 2009; Makris et al., 2009). The MdLF is a fiber bundle that runs within the white matter of the superior temporal gyrus extending from the IPL to the temporal pole. Although there is no common agreement yet, it appears as if the EmC corresponds with the bundle termed "inferior occipitofrontal fasciculus (IOF)" [also "inferior frontooccipital fasciculus (IFOF)"] by other authors (Nieuwenhuys, Voogd, & van Huijzen, 1988; Catani et al., 2002; Kier, Staib, Davis, & Bronen, 2004; Wakana, Jiang, Nagae-Poetscher, van Zijl, & Mori, 2004; Bürgel et al., 2006).

For clarification it should be pointed out that Catani and colleagues used a different terminology when they investigated the long perisylvian association fibers of the human left hemisphere (Catani et al., 2002, 2005, 2007). The SLF and AF historically have been regarded as a single fiber bundle in the human (Burdach, 1819–26; Dejerine & Dejerine-Klumpke, 1895). The terms "superior longitudinal fasciculus" and "arcuate fasciculus" thus often were and still are used interchangeably by some authors, including Catani and coworkers. However, despite the different terminology, Catani and colleagues (2005) also found a long and two shorter segments between the superior/middle temporal, inferior parietal, and lateral frontal cortices (figure 17.1*A*). Tractography reconstruction for a group of 11 healthy subjects revealed a direct connection between the left rentrolateral frontal and the superior/middle temporal cortex—that is, between Broca's and Wernicke's language areas. In addition, two shorter pathways were found connecting superior/middle temporal with the inferior parietal

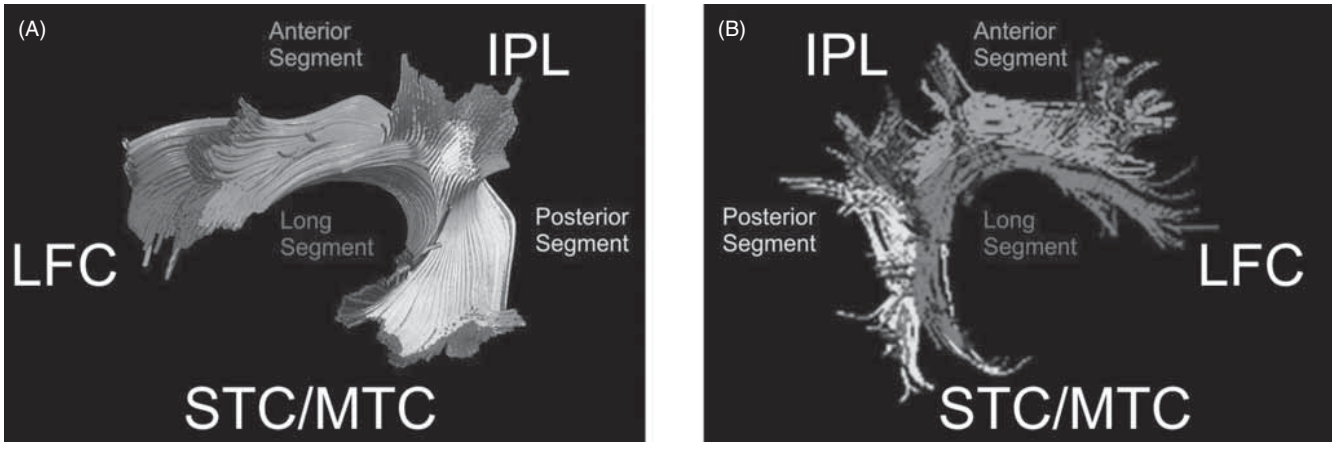


Figure 17.1 Averaged tractography reconstruction for fiber connections between the superior/middle temporal, inferior parietal, and lateral frontal cortices by using a two-region-of-interest approach in (*A*) the human left hemisphere (Catani, Jones, & ffytche, 2005) and (*B*) the human right hemisphere (Gharabaghi et al., 2009). A long connection was observed linking superior/middle temporal and lateral frontal cortices (shown in red). Two shorter pathways also were found. The posterior segment running from the superior/middle temporal to the inferior parietal cortex is shown in yellow. The anterior segment running from the inferior parietal to the lateral frontal cortex is shown in green. IPL, inferior parietal lobule; LFC, lateral frontal cortex; STC, superior temporal cortex; MTC, middle temporal cortex. (With modifications from Catani et al., 2005, and from Gharabaghi et al., 2009.) (See color plate 19.)

cortex ("posterior segment") and the inferior parietal with the dorsolateral frontal cortex ("anterior segment"). There is no doubt that the long connection between the superior/middle temporal and inferior frontal cortex represents the fiber bundle that has been termed AF in the work of other groups (Petrides & Pandya, 1988; Makris et al., 2005; Schmahmann et al., 2007; Upadhyay et al., 2008). The "anterior segment" between the inferior parietal and inferior frontal cortex most probably represents the fiber bundle(s) that have been termed SLF II—maybe in combination with the SLF III and/or SOF. The "posterior segment" between the superior temporal and inferior parietal cortex had been assumed to represent the MdLF (Schmahmann et al., 2007). However, the recent work by Makris et al. (2009) in the human rather argues that the MdLF is distinct from and located medial to the SLF-AF fibers.

To analyze the perisylvian connectivity between the superior/middle temporal, inferior parietal, and lateral frontal cortices in the human right hemisphere, Gharabaghi and coworkers (2009) investigated 12 right-handed male subjects without neurological deficits by using the same procedure that Catani and colleagues (2005) applied for the left hemisphere analysis. Figure 17.1*B* shows the averaged tractography reconstruction obtained from this DTI analysis. It revealed a pattern of fiber connections that largely corresponded to the one demonstrated by Catani and colleagues (2005) in the human left hemisphere (figure 17.1*A*). While Gharabaghi and coworkers were conducting this analysis, Catani and colleagues published a study (Catani et al., 2007) in which they also had analyzed the perisylvian connectivity in the human right hemisphere. In line with the findings illustrated in figure 17.1*B*, they found an indirect connection with a posterior segment connecting the superior/middle temporal with the inferior parietal cortex and an anterior segment running from the inferior parietal to the dorsolateral frontal cortex. In contrast, they found the long, direct segment between the superior/middle temporal and the lateral frontal cortices in only about 40% of their individuals, while this segment was present in all subjects (100%) studied by Gharabaghi and colleagues (2009). Likewise, some studies observed largely symmetrical conditions between the human hemispheres for volume, bundle density, and location of the left- and right-sided AF and SLF (Makris et al., 2005; Bürgel et al., 2006; Upadhyay et al., 2008), while discrepant observations have also been reported (Powell et al., 2006; Vernooij et al., 2007; Glasser & Rilling, 2008). Possible reasons for the discrepancy between these studies can be attributed to differences in fiber tracking methods, in the choice of the seeding ROIs, and/or the composition of subject samples. Future studies will have to clarify this issue. However, beyond the discrepant observations regarding the long, dorsally located direct connection via the AF, it is undisputed that the superior/middle temporal, lateral frontal, and inferior parietal cortices show dense direct (via the EmC/IOF) as well as indirect interconnectivity.

To summarize the hitherto existing findings from tract-tracing, myelin-staining, and diffusion-based imaging techniques, a dense perisylvian network seems to exist in both hemispheres connecting the inferior parietal lobule with the ventrolateral frontal cortex (via SLF II, SLF III, SOF), ventrolateral frontal cortex with superior/middle temporal cortex (via AF, EmC/IOF), and superior temporal cortex with the inferior parietal lobule (via MdLF, EmC/IOF). Figure 17.2 illustrates these tightly connected perisylvian neural networks.

FIGURE 17.2 Sketch of the perisylvian neural network linking the inferior parietal lobule with the ventrolateral frontal cortex (via SLF II, SLF III, SOF), ventrolateral frontal cortex with superior/middle temporal cortex and insula (via AF, EmC/IOF), and superior temporal cortex with the inferior parietal lobule (via MdLF, EmC/IOF). SLF II/III, subcomponents II/III of the superior longitudinal fasciculus; SOF, superior occipitofrontal fasciculus; AF, arcuate fasciculus; IOF, inferior occipitofrontal fasciculus; EmC, extreme capsule; MdLF, middle longitudinal fasciculus.

Functional role of the perisylvian network in the human right hemisphere

There is no disagreement that the perisylvian network in the human left hemisphere is involved in language processes (e.g., Frey, Campbell, Pike, & Petrides, 2008; Saur et al., 2008; Catani & Mesulam, 2008; Makris & Pandya, 2009). In contrast, the functional involvement of the right hemisphere perisylvian network is less clear. Catani and colleagues suggested that the perisylvian network in the human right hemisphere might represent—as in the human left hemisphere—a network involved in language functions (Catani et al., 2007, p. 17166). In this chapter, a different view is suggested. The perisylvian pathways between the right IPL, ventrolateral frontal, and superior/middle temporal cortices and insula connect those areas which have repeatedly been associated with spatial neglect in the case of brain damage (Heilman et al., 1983; Vallar & Perani, 1986; Mort et al., 2003; Karnath et al., 2001, 2004; Committeri et al., 2007; Sarri et al., 2009). In contrast, aphasia is only extremely rarely associated with lesion of these right hemisphere perisylvian areas (as rarely as spatial neglect is observed after left hemisphere damage). Thus it is proposed that the perisylvian network in the human right hemisphere represents the anatomical basis for processes involved in spatial orienting and exploration.

Supporting evidence for this hypothesis has been reported from transcranial magnetic stimulation (TMS), electrical mapping of the human cortex during neurosurgery, and fMRI in healthy subjects. Using TMS, Ellison, Schindler, Pattison, and Milner (2004) induced "virtual lesions" at the right STG and right posterior parietal cortex (PPC) in healthy subjects. They observed a specific impairment induced by TMS over the right STG for serial feature search (termed "hard feature search task"). In contrast, TMS over the right PPC resulted in increased reaction times during "hard conjunction search." Gharabaghi, Fruhmann-Berger, Tatagiba, and Karnath (2006) observed that intraoperative inactivation of the middle portion of the STG in human leads to disturbed serial visual search. Using the same technique Thiebaut de Schotten and colleagues (2005) found that inactivating regions in the right IPL or at the caudal and the middle parts of the STG leads to deficits in the perception of line length.

Evidence for the involvement of superior temporal, inferior parietal, and lateral frontal areas in processes of spatial orienting has also been obtained from fMRI experiments in healthy subjects. In a cued spatial-attention task, Hopfinger, Buonocore, and Mangun (2000) found bilateral activation in these cortical areas correlated with covert attentional shifts in the horizontal dimension of space. Himmelbach, Erb, and Karnath (2006) investigated active visual exploration in healthy subjects, using a task (visual search in a letter array) that closely resembled the clinical procedure employed to detect spatial neglect in stroke patients. The authors observed significant activation associated with visual exploration located at the TPJ, the midportion of the STG, and the IFG.

Thus observations deriving from different techniques converge to suggest that the densely interconnected perisylvian neural system in the human right hemisphere (figure 17.2) represents the anatomical basis of processes that are involved in spatial orientation, provoking spatial neglect in the case of damage. The tight anatomical connectivity between superior/middle temporal, inferior parietal, and ventrolateral frontal cortices might explain why lesions at these distant cortical sites around the sylvian fissure in the human right hemisphere can lead to the same disturbance of orienting behavior, namely, to spatial neglect.

Spatial neglect—A disconnection syndrome?

Beginning with the seminal work of Dejerine and his wife (Dejerine & Dejerine-Klumpke, 1895) on the human cortical pathways, several authors developed the idea that some neurological conditions might result from the disconnection of one area of the brain from another. Among them, Geschwind (1965) put forward the view that several neuropsychological disorders could best be interpreted as resulting from interruption of specific cortical association pathways. With respect to spatial neglect, Mesulam and Geschwind (1978) suggested that this disorder—among other disorders of attention and emotion—results from disruption of neural connections between limbic structures and neocortex. Mesulam (1981, 1985) further evolved this concept, suggesting that an interconnected network between posterior parietal, frontal, and cingulate cortices as well as the reticular formation is involved in spatial neglect. A disconnection hypothesis has also been put forward by Watson, Heilman, Miller, and King (1974) and Watson, Miller, and Heilman (1978) when observing that spatial neglect can be evoked in the monkey by a lesion in the mesencephalic reticular formation.

More recently, some authors have revived the concept to view spatial neglect as a "disconnection syndrome" (Catani, 2006; Bartolomeo, Thiebaut de Schotten, & Doricchi, 2007; He et al., 2007). Bartolomeo and colleagues proposed that long-lasting signs of spatial neglect result from frontoparietal intrahemispheric and from interhemispheric disconnection. They suggested that "a particular form of disconnection might have greater predictive value than the localization of gray matter lesions concerning the patients' deficits and disabilities" (Bartolomeo et al., 2007, p. 2484). Intrahemispherically, they related disconnection of the SLF (Thiebaut de Schotten et al., 2005, 2008; Bartolomeo et al., 2007) but also of the IOF (Urbanski et al., 2008) to spatial neglect. Using DTI tractography, He and colleagues found damage to the

SLF and AF in five patients with severe spatial neglect but not in five patients with mild cases. Furthermore, the analysis of interregional functional connectivity, based on coherent fluctuations of fMRI signals, suggested that not only anatomically but also functionally disrupted connectivity in dorsal and ventral attention networks might constitute a critical mechanism underlying the pathophysiology of spatial neglect (He et al., 2007).

To investigate the possible impact of damage to white matter association fibers for the genesis of spatial neglect, Karnath, Rorden, and Ticini (2009) analyzed lesion location in a large seven-year sample of 140 right-hemispheric stroke patients. This large number of stroke patients allowed the authors not only to study a representative sample of subjects with spatial neglect, but also to perform a statistical voxel-wise lesion-behavior mapping (VLBM) analysis (e.g., Bates et al., 2003; Rorden, Karnath, & Bonilha, 2007) to estimate which brain regions are more frequently compromised in neglect patients relative to patients without neglect. Karnath and coworkers (2009) studied the patients' white matter connectivity by using a new method that combines a statistical VLBM approach with the histological maps of the human white matter fiber tracts provided by the stereotaxic probabilistic atlas developed by the Jülich group (Amunts & Zilles, 2001; Zilles, Schleicher, Palomero-Gallagher, & Amunts, 2002). In contrast to the reference brain of the Talairach and Tournoux atlas (Talairach & Tournoux, 1988) or the MNI single-subject or group templates (Evans et al., 1992; Collins, Neelin, Peters, & Evans, 1994), the Jülich probabilistic atlas is based on the analysis of the cytoarchitecture in a sample of 10 different human postmortem brains. It thus provides information on the location and intersubject variability of brain structures, illustrating for each voxel of the MNI reference space the relative frequency with which a certain structure was present in 10 normal human brains. Using a modified myelin-staining technique, Bürgel and colleagues (2006) were able to distinguish 10 individual white matter fiber tracts for this atlas at microscopic resolution.

The analysis of the 140 right-hemisphere stroke patients revealed that 7.0% of the right SLF, 8.2% of the IOF, 12.7% of the SOF, and only 0.6% of the uncinate fasciculus were significantly more affected in patients with spatial neglect than in those not showing the disorder (figure 17.3). The authors concluded that damage of right perisylvian white matter connections is a typical finding in patients with spatial neglect. However, the proportion of involvement of each of the fiber bundles was very low. When the authors analyzed how much of the lesion area in neglect patients overlapped with all of the perisylvian white matter connections, they found an overlap between 3.4% and 10.9% (Karnath et al., 2009).

Although the study by Karnath and coworkers (2009) cannot finally decide whether or not spatial neglect should best be interpreted as a "disconnection syndrome" (Mesulam & Geschwind, 1978; Watson et al., 1974; Watson, Miller, & Heilman, 1978; Catani, 2006; Bartolomeo et al., 2007; He et al., 2007), one may conclude that their data argue more against than in favor of such a hypothesis. In fact, their analysis revealed that between 89.1% and 96.6% of the lesion area in spatial neglect affected brain structures other than the perisylvian white matter fiber tracts, namely, cortical and subcortical gray matter structures such as the superior temporal, inferior parietal, inferior frontal, and insular cortices, as well as the putamen and caudate nucleus (Karnath et al., 2009). Damage to these gray matter structures in the right hemisphere thus appears to be a strong predictor of spatial neglect.

Another aspect arguing against the view of spatial neglect as a white matter disconnection syndrome in the traditional sense is the perfusion-weighted imaging (PWI) results obtained in patients with subcortical infarcts. Perfusion-weighted imaging is an MR technique that allows the identification of brain regions that are receiving enough blood supply to remain structurally intact but not enough to function normally. By using this technique, several studies showed that left- or right-sided subcortical lesions—including selective white matter strokes—cause spatial neglect only if the subcortical damage provokes additional malperfusion of cortical gray matter structures in the ipsilesional hemisphere (Demeurisse, Hublet, Paternot, Colson, & Serniclaes, 1997; Hillis et al., 2002, 2005). Without this malfunction of cortical structures, subcortical brain lesions did not provoke disturbances of spatial orienting. Thus it seems that damage to subcortical white matter connectivity alone does not provoke spatial neglect but rather requires additional malfunction of cortical gray matter structures.

Conclusions

Homologous perisylvian neural networks seem to exist in the human left and right hemispheres composed of tightly connected cortical areas straddling the sylvian fissure (cf. figure 17.2). It is suggested that the neural network consisting of superior/middle temporal, inferior parietal, and ventrolateral frontal cortices in the human right hemisphere represents the anatomical basis for processes involved in spatial orienting. Neurons of these regions provide us with redundant information about the position and motion of our body in space. They seem to play an essential role in adjusting body position relative to external space (Karnath & Dieterich, 2006). Damage to this perisylvian system in the right hemisphere may provoke spatial neglect. In the human left hemisphere, a similar perisylvian network seems to exist but is serving different functions, namely, language and praxis. This functional specialization of left and right perisylvian networks is still not observed in the nonhuman

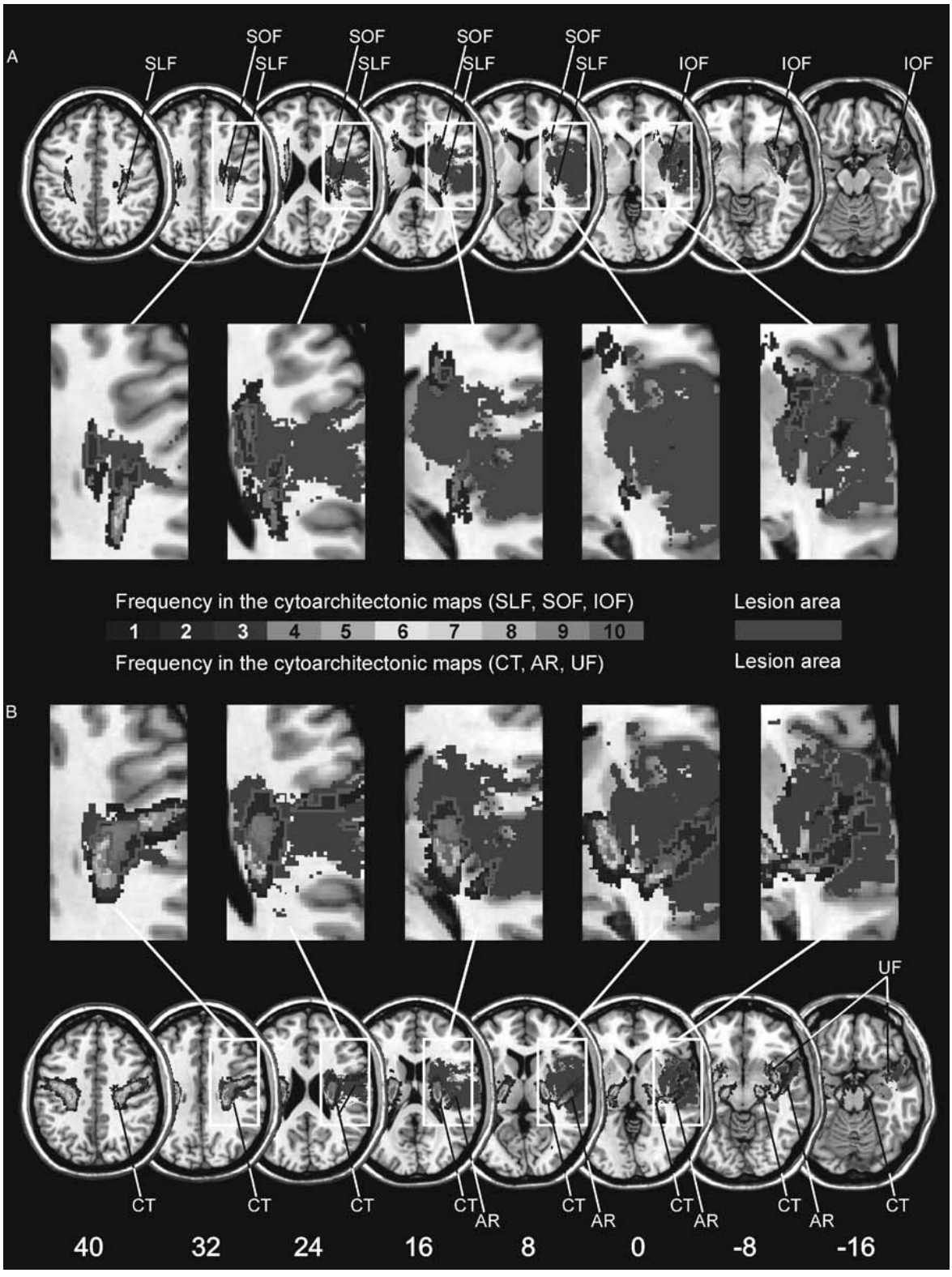


FIGURE 17.3 Overlap of the statistical VLBM lesion map (the brain territory significantly more affected in 78 patients with spatial neglect than in 62 stroke patients without this disorder) with the probabilistic, cytoarchitectonic maps of the white matter association fiber tracts from the Jülich atlas. The statistical lesion map is illustrated in homogeneous brown color. The color coding of the Jülich atlas from 1 (dark blue, observed in 1 postmortem brain) to 10 (red, overlap in all ten postmortem brains) represents the absolute frequency for which in each voxel of the brain a respective fiber tract was present (e.g., yellow color indicates that the fiber tract was present in that voxel in seven out of ten postmortem brains). The pink contour demarks the area of the fiber tracts affected by the statistical lesion map. (*A*) Overlap illustrated for perisylvian fiber tracts SFL, superior longitudinal fasciculus; IOF, inferior occipitofrontal fasciculus; and SOF, superior occipitofrontal fasciculus. (*B*) Overlap illustrated for fiber tracts CT, corticospinal tract; AR, acoustic radiation; and UF, uncinate fascicle. (From Karnath et al., 2009.) (See color plate 20.)

primate. Here, lesions of this perisylvian system in both hemispheres induce disturbed exploration and orientation toward the respective contralateral side (e.g., Luh, Butter, & Buchtel, 1986; Watson, Valenstein, Day, & Heilman, 1994; Wardak, Olivier, & Duhamel, 2002, 2004). Hence the phylogenetic transition from monkey to human brain seems to be a restriction of a formerly bilateral function represented within right- and left-sided perisylvian networks to the right hemisphere (Karnath et al., 2001). It appears as if this lateralization of spatial orientation to the right hemisphere network parallels the emergence of an elaborate representation for language in the left-sided perisylvian network.

ACKNOWLEDGMENTS This work was supported by the Bundesministerium für Bildung und Forschung (BMBF-Verbundprojekt "Räumliche Orientierung" 01GW0641) and the Deutsche Forschungsgemeinschaft (SFB 550-A4). I would like to thank Bianca de Haan and Marc Himmelbach for their discussion and helpful comments on the manuscript.

REFERENCES

Amunts, K., & Zilles, K. (2001). Advances in cytoarchitectonic mapping of the human cerebral cortex. *Neuroimaging Clin. N. Am.*, *11*, 151–169.

Bartolomeo, P., Thiebaut de Schotten, M., & Doricchi, F. (2007). Left unilateral neglect as a disconnection syndrome. *Cereb. Cortex*, *17*, 2479–2490.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion–symptom mapping. *Nat. Neurosci.*, *6*, 448–450.

Behrmann, M., Watt, S., Black, S. E., & Barton, J. J. (1997). Impaired visual search in patients with unilateral neglect: An oculographic analysis. *Neuropsychologia*, *35*, 1445–1458.

Boatman, D. (2004). Cortical bases of speech perception: Evidence from functional lesion studies. *Cognition*, *92*, 47–65.

Borovsky, A., Saygin, A. P., Bates, E., & Dronkers, N. (2007). Lesion correlates of conversational speech production deficits. *Neuropsychologia*, *45*, 2525–2533.

Burdach, K. F. (1819–1826). *Vom Baue und Leben des Gehirns*. Leipzig: Dyk.

Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J. M., & Zilles, K. (2006). White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability. *NeuroImage*, *29*, 1092–1105.

Buxbaum, L. J., Ferraro, M. K., Veramonti, T., Farne, A., Whyte, J., Ladavas, E., et al. (2004). Hemispatial neglect: Subtypes, neuroanatomy, and disability. *Neurology*, *62*, 749–756.

Catani, M. (2006). Diffusion tensor magnetic resonance imaging tractography in cognitive disorders. *Curr. Opin. Neurol.*, *19*, 599–606.

Catani, M., Allin, M. P. G., Husain, M., Pugliese, L., Mesulam, M.-M., Murray, R. M., et al. (2007). Symmetries in human brain language pathways correlate with verbal recall. *Proc. Natl. Acad. Sci. USA*, *104*, 17163–17168.

Catani, M., Howard, R. J., Pajevic, S., & Jones, D. K. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage*, *17*, 77–94.

Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.*, *57*, 8–16.

Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state. *Cortex*, 44, 953–961.

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.*, *18*, 192–205.

Committeri, G., Pitzalis, S., Galati, G., Patria, F., Pelle, G., Sabatini, U., et al. (2007). Neural bases of personal and extrapersonal neglect in humans. *Brain*, *130*, 431–441.

Corbetta, M., Kincade, M. J., Lewis, C., Snyder, A. Z., & Sapir, A. (2005). Neural basis and recovery of spatial attention deficits in spatial neglect. *Nat. Neurosci.*, *8*, 1603–1610.

Dejerine, J., & Dejerine-Klumpke, A. M. (1895). *Anatomie des centres nerveux*. Paris: Rueff et Cie.

Demeurisse, G., Hublet, C., Paternot, J., Colson, C., & Serniclaes, W. (1997). Pathogenesis of subcortical visuospatial neglect: A HMPAO SPECT study. *Neuropsychologia*, *35*, 731–735.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Jr., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*, 145–177.

Duffau, H., Gatigno, P., Mandonnet, E., Peruzzi, P., Tzourio-Mazoyer, N., & Capelle, L. (2005). New insights into the anatomo-functional connectivity of the semantic system: A study using cortico-subcortical electrostimulations. *Brain*, *128*, 797–810.

Ellison, A., Schindler, I., Pattison, L. L., & Milner, A. D. (2004). An exploration of the role of the superior temporal gyrus in visual search and spatial perception using TMS. *Brain*, *127*, 2307–2315.

Evans, A. C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W., et al. (1992). Anatomical mapping of functional activation in stereotactic coordinate space. *NeuroImage*, *1*, 43–53.

Frey, S., Campbell, J. S., Pike, G. B., & Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.*, *28*, 11435–11444.

Fruhmann-Berger, M., & Karnath, H.-O. (2005). Spontaneous eye and head position in patients with spatial neglect. *J. Neurol.*, *252*, 1194–1200.

Fruhmann-Berger, M., Pross, R. D., Ilg, U. J., & Karnath, H.-O. (2006). Deviation of eyes and head in acute cerebral stroke. *BMC Neurol.*, *6*, 23; corrigendum, *6*, 49.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain*, *88*, 237–294, 585–644.

Gharabaghi, A., Fruhmann-Berger, M., Tatagiba, M., & Karnath, H.-O. (2006). The role of the right superior temporal gyrus in visual search—Insights from intraoperative electrical stimulation. *Neuropsychologia*, *44*, 2578–2581; corrigendum, *45*, 465.

Gharabaghi, A., Kunath, F., Erb, M., Saur, R., Heckl, S., Tatagiba, M., et al. (2009). Perisylvian white matter connectivity in the human right hemisphere. *BMC Neurosci.*, *10*, 15.

Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cereb. Cortex*, *18*, 2471–2482.

Goldenberg, G., HermsdÖrfer, J., Glindemann, R., Rorden, C., & Karnath, H.-O. (2007). Pantomime of tool use depends on integrity of left inferior frontal cortex. *Cereb. Cortex*, *17*, 2769–2776.

Goldenberg, G., & Karnath, H.-O. (2006). The neural basis of imitation is body part specific. *J. Neurosci.*, *26*, 6282–6287.

He, B. J., Snyder, A. Z., Vincent, J. L., Epstein, A., Shulman, G. L., & Corbetta, M. (2007). Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron*, *53*, 905–918.

Heilman, K. M., Watson, R. T., Valenstein, E., & Damasio, A. R. (1983). Localization of lesions in neglect. In A. Kertesz (Ed.), *Localization in neuropsychology* (pp. 471–492). New York: Academic Press.

Hillis, A. E., Newhart, M., Heidler, J., Barker, P. B., Herskovits, E. H., & Degaonkar, M. (2005). Anatomy of spatial attention: Insights from perfusion imaging and hemispatial neglect in acute stroke. *J. Neurosci.*, *25*, 3161–3167.

Hillis, A. E., Wityk, R. J., Barker, P. B., Beauchamp, N. J., Gailloud, P., Murphy, K., et al. (2002). Subcortical aphasia and neglect in acute stroke: The role of cortical hypoperfusion. *Brain*, *125*, 1094–1104.

Himmelbach, M., Erb, M., & Karnath, H.-O. (2006). Exploring the visual world: The neural substrate of spatial orienting. *NeuroImage*, *32*, 1747–1759.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat. Neurosci.*, *3*, 284–291.

Husain, M., & Kennard, C. (1996). Visual neglect associated with frontal lobe infarction. *J. Neurol.*, *243*, 652–657.

Iacoboni, M., & Wilson, S. M. (2006). Beyond a single area: Motor control and language within a neural architecture encompassing Broca's area. *Cortex*, *42*, 503–506.

Karnath, H.-O., & Dieterich, M. (2006). Spatial neglect—a vestibular disorder? *Brain*, *129*, 293–305.

Karnath, H.-O., Ferber, S., & Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*, *411*, 950–953.

Karnath, H.-O., Fruhmann-Berger, M., Küker, W., & Rorden, C. (2004). The anatomy of spatial neglect based on voxelwise statistical analysis: A study of 140 patients. *Cereb. Cortex*, *14*, 1164–1172.

Karnath, H.-O., Niemeier, M., & Dichgans, J. (1998). Space exploration in neglect. *Brain*, *121*, 2357–2367.

Karnath, H.-O., Rorden, C., & Ticini, L. F. (2009). Damage to white matter fiber tracts in acute spatial neglect. *Cereb. Cortex*, in press.

Kertesz, A., Harlock, W., & Coates, R. (1979). Computer tomographic localization, lesion size, and prognosis in aphasia and nonverbal impairment. *Brain Lang.*, *8*, 34–50.

Kier, E. L., Staib, L. H., Davis, L. M., & Bronen, R. A. (2004). MR imaging of the temporal stem: Anatomic dissection tractography of the uncinate fasciculus, inferior occipitofrontal fasciculus, and Meyer's Loop of the optic radiation. *Am. J. Neuroradiol.*, *25*, 677–691.

Kreisler, A., Godefroy, O., Delmaire, C., Debachy, B., Leclercq, M., Pruvo, J.-P., et al. (2000). The anatomy of aphasia revisited. *Neurology*, *54*, 1117–1123.

Luh, K. E., Butter, C. M., & Buchtel, H. A. (1986). Impairments in orienting to visual stimuli in monkeys following unilateral lesions of the superior sulcal polysensory cortex. *Neuropsychologia*, *24*, 461–470.

Makris, N., Kennedy, D. N., McInerney, S., Sorensen, A. G., Wang, R., Caviness, V. S., Jr., et al. (2005). Segmentation of subcomponents within the superior longitudinal fascicle in humans: A quantitative, in vivo, DT-MRI study. *Cereb. Cortex*, *15*, 854–869.

Makris, N., & Pandya, D. N. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Struct. Funct.*, *213*, 343–358.

Makris, N., Papadimitriou, G. M., Kaiser, J. R., Sorg, S., Kennedy, D. N., & Pandya, D. N. (2009). Delineation of the middle longitudinal fascicle in humans: A quantitative, in vivo, DT-MRI study. *Cereb. Cortex*, *19*, 777–785.

Makris, N., Papadimitriou, G. M., Sorg, S., Kennedy, D. N., Caviness, V. S., & Pandya, D. N. (2007). The occipitofrontal fascicle in humans: A quantitative, in vivo, DT-MRI study. *NeuroImage*, *37*, 1100–1111.

Mesulam, M.-M. (1981). A cortical network to directed attention and unilateral neglect. *Ann. Neurol.*, *10*, 309–325.

Mesulam, M.-M. (1985). Attention, confusional states, and neglect. In M.-M. Mesulam (Ed.), *Principles of behavioral neurology* (pp. 125–168). Philadelphia: F. A. Davis.

Mesulam, M.-M., & Geschwind, N. (1978). On the possible role of neocortex and its limbic connections in the process of attention and schizophrenia: Clinical cases of inattention in man and experimental anatomy in monkey. *J. Psychiatr. Res.*, *14*, 249–259.

Mori, S., Wakana, S., van Zijl, P. C. M., & Nagae-Poetscher, L. M. (2005). *MRI atlas of human white matter*. Amsterdam: Elsevier.

Mort, D. J., Malhotra, P., Mannan, S. K., Rorden, C., Pambakian, A., Kennard, C., et al. (2003). The anatomy of visual neglect. *Brain*, *126*, 1986–1997.

Nieuwenhuys, R., Voogd, J., & van Huijzen, C. (1988). *The human central nervous system*. Berlin: Springer.

Petrides, M., & Pandya, D. N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.*, *228*, 105–116.

Petrides, M., & Pandya, D. N. (1988). Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J. Comp. Neurol.*, *273*, 52–66.

Petrides, M., & Pandya, D. N. (2007). Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *J. Neurosci.*, *27*, 11573–11586.

Poeck, K., de Bleser, R., & Graf von Keyserlingk, D. (1984). Computed tomography localization of standard aphasic syndromes. In F. C. Rose (Ed.), *Advances in neurology*, Vol. 42: *Progress in aphasiology* (pp. 71–89). New York: Raven Press.

Powell, H. W. R., Parker, G. J. M., Alexander, D. C., Symms, M. R., Boulby, P. A., Wheeler-Kingshott, C. A. M., et al. (2006). Hemispheric asymmetries in language-related pathways: A combined functional MRI and tractography study. *NeuroImage*, *32*, 388–399.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.*, *21*, 188–194.

Rorden, C., Karnath, H.-O., & Bonilha, L. (2007). Improving lesion-symptom mapping. *J. Cogn. Neurosci.*, *19*, 1081–1088.

Sanai, N., Mirzadeh, Z., & Berger, M. S. (2008). Functional outcome after language mapping for glioma resection. *N. Engl. J. Med., 358*, 18–27.

Sarri, M., Greenwood, R., Kalra, L., & Driver, J. (2009). Task-related modulation of visual neglect in cancellation tasks. *Neuropsychologia*, 47, 91–103.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M. S., et al. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. USA*, *105*, 18035–18040.

Schmahmann, J. D., & Pandya, D. N. (2006). *Fiber pathways of the brain*. New York: Oxford University Press.

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., et al. (2007). Association fiber pathways of the brain: Parallel observations from diffusion spectrum imaging and autoradiography. *Brain*, *130*, 630–653.

Seltzer, B., & Pandya, D. N. (1984). Further observations on parieto-temporal connections in the rhesus monkey. *Exp. Brain Res.*, *55*, 301–312.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. Stuttgart: Thieme.

Thiebaut de Schotten, M., Kinkingnéhun, S., Delmaire, C., Lehéricy, S., Duffau, H., Thivard, L., et al. (2008). Visualization of disconnection syndromes in humans. *Cortex*, *44*, 1097–1103.

Thiebaut de Schotten, M., Urbanski, M., Duffau, H., Volle, E., Lévy, R., Dubois, B., et al. (2005). Direct evidence for a parietal-frontal pathway subserving spatial awareness in humans. *Science*, *309*, 2226–2228.

Upadhyay, J., Hallock, K., Ducros, M., Kim, D.-S., & Ronen, I. (2008). Diffusion tensor spectroscopy and imaging of the arcuate fasciculus. *NeuroImage*, *39*, 1–9.

Urbanski, M., Thiebaut de Schotten, M., Rodrigo, S., Catani, M., Oppenheim, C., Touzé, E., et al. (2008). Brain networks of spatial awareness: Evidence from diffusion tensor imaging tractography. *J. Neurol. Neurosurg. Psychiatry*, *79*, 598–601.

Vallar, G., & Perani, D. (1986). The anatomy of unilateral neglect after right-hemisphere stroke lesions: A clinical/CT-scan correlation study in man. *Neuropsychologia*, *24*, 609–622.

Vernooij, M. W., Smits, M., Wielopolski, P. A., Houston, G. C., Krestin, G. P., & van der Lugt, A. (2007). Fiber density asymmetry of the arcuate fasciculus in relation to functional hemispheric language lateralization in both right- and left-handed healthy subjects: A combined fMRI and DTI study. *NeuroImage*, *35*, 1064–1076.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., et al. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*, 1414—1432.

Wakana S., Jiang, H., Nagae-Poetscher, L. M., van Zijl, P. C. M., & Mori, S. (2004). Fiber tract-based atlas of human white matter anatomy. *Radiology*, *230*, 77–87.

Wardak, C., Olivier, E., & Duhamel, J.-R. (2002). Neglect in monkeys: Effect of permanent and reversible lesions. In H.-O. Karnath, A. D. Milner, & G. Vallar (Eds.), *The cognitive and neural bases of spatial neglect* (pp. 101–118). Oxford, UK: Oxford University Press.

Wardak, C., Olivier, E., & Duhamel, J.-R. (2004). A deficit in covert attention after parietal cortex inactivation in the monkey. *Neuron*, *42*, 501–508.

Watson, R. T., Heilman, K. M., Miller, B. D., & King, F. A. (1974). Neglect after mesencephalic reticular formation lesions. *Neurology*, *24*, 294–298.

Watson, R. T., Miller, B. D., & Heilman, K. M. (1978). Nonsensory neglect. *Ann. Neurol.*, *3*, 505–508.

Watson, R. T., Valenstein, E., Day, A., & Heilman, K. M. (1994). Posterior neocortical systems subserving awareness and neglect. *Arch. Neurol.*, *51*, 1014–1021.

Zilles, K., Schleicher, A., Palomero-Gallagher, N., & Amunts, K. (2002). Quantitative analysis of cyto- and receptor architecture of the human brain. In J. C. Mazziotta & A. Toga (Eds.), *Brain mapping: The methods* (pp. 573– 602). Amsterdam: Elsevier.

18 Spatial Deficits and Selective Attention

LYNN C. ROBERTSON

ABSTRACT The focus of this chapter is on spatial deficits that produce a complete or partial loss of spatial awareness of the visual world after damage to the dorsal pathway of the human brain. Not surprisingly, when spatial awareness is deficient, controlling spatial attention is also compromised. Yet even with complete loss of spatial information of the external world, object-based and feature-based processes continue to influence what is seen. Nevertheless, features may be inaccurately bound to form abnormal rates of illusory conjunctions even under free viewing conditions. There also is emerging evidence from priming studies that conjunctions are bound late in processing whereas features are coded early. How the multiple spatial representations in the brain may interact to influence selection is also discussed.

The influential 18th-century philosopher Immanuel Kant claimed that space and time were the two necessary mental concepts supporting all other human experience. A mental representation of space separates sensory experience occurring at the same time into different entities that are spatially segregated yet related to one another, while a mental representation of time separates sequentially presented information into segmented events. Kant himself wrote, "We never can imagine or make a representation to ourselves of the nonexistence of space."

Although it is very nearly impossible to imagine a world in which space does not exist, there are individuals with damage to certain brain areas who must contend with the loss of spatial perception on a daily basis. These are neurological patients who have suffered unilateral damage to parietal (and/or less often frontal or superior temporal) areas, producing unilateral neglect, and those with bilateral parietal damage, which can produce a complete loss of spatial information beyond a person's own body (Balint's syndrome). Studies of such individuals have shown that certain perceptual experiences remain relatively intact, but others are altered or lost altogether. Contrary to Kant's claims, a mental representation of space is not necessary for all perceptual phenomena, and the exceptions provide insights into the cognitive and neurobiological bases of spatial representations and their interactions with perception and other attention mechanisms.

LYNN C. ROBERTSON Veterans Administration Research, Department of Psychology, and Helen Wills Neuroscience Institute, University of California, Berkeley, California

The loss of perceptual space

I will begin by classifying spatial deficits in behavioral neurology into three general classes: complete (there is no there there), partial (only a portion is there), and scrambled (here when it should be there). Complete, or nearly complete, loss of a mental spatial map can be observed in Balint's syndrome (Balint, 1909; Holmes & Horax, 1919; Rafal, 1997), partial loss can be observed in unilateral neglect (Heilman, Watson, & Valenstein, 1994; Bartolomea & Chokron, 2001), and scrambling is seen in integrative agnosia, where stimuli are processed piecemeal, producing a fragmented percept of parts with little overall coherence (Riddoch & Humphreys, 1987a). Although any of these can be observed in more than one modality, the deficits appear most often or at least are more obvious in vision.

All three types of spatial loss have been associated with (although not necessarily limited to) posterior damage of the human brain, with complete and partial loss more prevalent after dorsal damage (although see Karnath, Ferber, & Himmelback, 2001), whereas spatial scrambling is more prevalent after ventral damage. Complete loss occurs after bilateral dorsal damage, and partial loss occurs after unilateral damage. Also, partial loss and scrambling are more likely to occur after right than eft hemisphere damage (Heilman et al., 1994; Ivry & Robertson, 1998; Mesulam, 1981). The common denominator for complete and partial loss is damage to the parietal lobe, but the areas involved are different depending on the nature of the spatial deficits. In fact, several researchers have suggested that the critical areas that produce unilateral neglect (a partial loss of space) are centered in the temporal-parietal junction and the inferior parietal lobe (e.g., Heilman et al., 1994; Mort et al., 2003, but see Karnath, chapter 17 in this volume). Conversely, a recent review of this literature has made a compelling argument for unilateral neglect as a disconnection syndrome. When lesions include the white matter tracts that connect posterior regions to the frontal lobe, neglect is more severe and more likely to be

chronic (Bartolomeo, Thiebaut de Schotten, & Doricchi, 2007).

Conversely, the common denominator in Balint's syndrome appears to be the angular gyrus and dorsal occipital association cortex bilaterally (Rizzo & Vecera, 2002; Rafal, 2001; L. Robertson, 2004). Although there are far fewer cases of patients with Balint's than with neglect, all cases of Balint's syndrome reported in the literature have included damage to these areas. The same areas were also implicated in a postmortem study by Hof, Bouras, Constantinidis, and Morrison (1989) examining the brains of a group of Alzheimer's patients who showed signs of Balint's syndrome during life (an occurrence rate of about 6%). He found more plaques and tangles in the dorsal occipital association cortex and angular gyrus (areas 19 and 39) than in any other portion of the visual brain. Accordingly, Balint's syndrome is not simply the bilateral version of unilateral neglect. Although lesions in unilateral neglect may extend into these areas on occasion, they are not the central areas implicated in the neglect syndrome.

Consequences of complete spatial loss and attention

Obviously, spatial attention will be disrupted if the spatial map that guides it is damaged or lost. When the computations needed to construct this spatial map are compromised, spatial attention deficits will be one major consequence. Thus Balint's syndrome presents an opportunity to study how deficits in spatial attention affect other perceptual and cognitive abilities. It also permits the study of how spatial maps support and interact with other perceptual phenomena. Intuitively, it seems that no meaningful percept would be possible without a spatial framework in which to individuate one entity from another or to guide selective attention within a visual display. Yet studies of cases of Balint's syndrome demonstrate that although the patients are functionally blind, they continue to experience a rather complex perceptual world in which features and objects rise to the level of perceptual awareness but their locations and spatial relationships to each other are unknown (Rafal, 2001). The result is a chaotic perceptual world, but not an absent one as Kant might maintain.

Balint's Syndrome and Object-Based Attention There is a great deal of evidence that attention can select objects and locations (see Treisman, chapter 12, this volume) but that these may be independent. In one of the first behavioral studies of object-based attention, Duncan (1984) showed that reporting values on two dimensions (orientation and texture) was faster when the two were on the same object rather than when they were on separate objects. Several other studies have shown that objects modulate spatial attention. For example, Egly, Driver, and Rafal (1994) presented two parallel rectangles on a screen and cued an end of one of the rectangles randomly on each trial (see Treisman, this volume, figure 12.2). The cue was followed by a target, and response time to detect the target was recorded. Importantly, the distance between the cue and the target in the same rectangle was the same as the distance between the cue and the target in the opposite rectangle. Responses were faster when the target appeared within the cued rectangle than when it appeared in the other rectangle (this difference is the object-based effect). Recently, Alice Albrecht, Alexandra List, and I (2008) showed that this object-based effect depends on whether the rectangles are perceived as objects or holes in a surface. We presented the same type of displays stereoscopically (figure 18.1), so that the rectangles were either seen as objects in front of a homogeneous surface or as two holes through that surface. We found object-based effects that were very similar in magnitude to that reported by Egly and colleagues, but only when the rectangles were perceived as objects. When the rectangles were perceived as holes, the object-based effect disappeared. Importantly these findings were not due to differences in perceived depth of the two rectangles. In a second study we split the background into individuated regions, and now the object-based effect reemerged (figures 18.1 and 18.2).

Object-based attention also has been studied in the fMRI literature. For instance, O'Craven, Downing, & Kanwisher (1999) showed two objects (a house and face) in the center

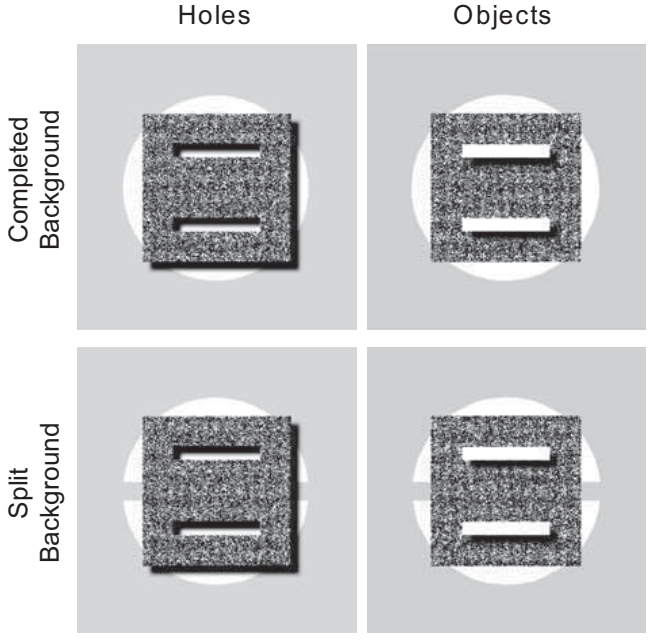
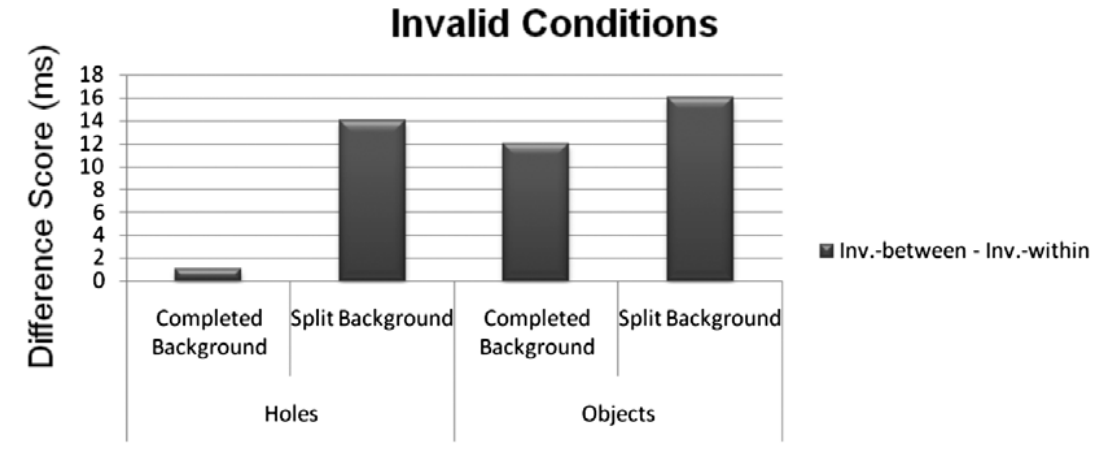


Figure 18.1 Example of hole and object stimuli with completed (top) and split (bottom) backgrounds. Shadows are included here only to illustrate the depth that was perceived by the participants when wearing the stereogoggles. They were not present in the actual experiment.



	Holes		Objects	
	Completed Background	Split Background	Completed Background	Split Background
Inv.-Within	422	409	422	403
Inv.-Between	423	423	434	419

FIGURE 18.2 Object-based effects (reaction time difference between the two invalid conditions—between minus within) for rectangles perceived as holes and those perceived as objects when the background was split and when it was completed (top). Mean reaction time for each invalid condition (bottom). The effects were significant for all but the completed background/holes condition.

of the screen. The two objects were superimposed and appeared transparent. On each trial either the house or the face moved slightly. While making judgments about the motion, there was more activity in ventral areas of the brain that respond to houses (parahippocampal place area) when the houses moved and more activity in the areas of the brain that respond to faces (fusiform face area) when faces moved. Attentional selection for motion (a feature that drives different areas of the brain than houses or faces) incorporated the object that was moving as well.

If object- and space-based attention utilize different neural systems, complete loss of space with Balint's syndrome should attract attention to objects, but there should be little or no knowledge of where they are located, and this result is what occurs. Individuals with this problem see only one object at any given moment (known as simultanagnosia), but they can be at chance in locating it. Neither reaching, pointing, nor verbally reporting where the object is located is accurate. Even spatial judgments such as saying whether the object is toward the top or bottom of the screen when they are several inches apart or toward the patient's own head or feet can be near chance levels. In addition, an object in a visual display will automatically and seemingly randomly attract attention with no control over how long the object remains in view or what object will take its place. The one object that is seen abruptly disappears and is replaced by another object (which need not be in the line of sight and can be large, small, simple, or complex). This new object is then perceived for a time, until another object takes its place. The computations needed to individuate objects or to define their locations are absent, and the objects that attract attention can come from anywhere in the visual scene. Although patients can move their eyes in any given direction on command, they only rarely do so unless instructed by the observer. They suffer from what Balint (1909) called a "pseudo paralysis of gaze." It is as if the one object they do see fills their entire visual field. Consistently, RM, the Balint's patient whom we studied for several years in my laboratory, reported that the size of objects did not appear as they should be.

The fact that an object as a whole can be perceived at all when spatial deficits are nearly complete is very strong evidence for an object-based system that is separate from a space-based one. However, when the spatial map that individuates objects is absent, voluntary movement of attention between objects is also a problem. It is as if there is a lineup of objects that compete for object selection when spatial information is gone.

BALINT'S SYNDROME AND FEATURE-BASED ATTENTION

Feature-based attention is also well supported in the perceptual literature. It too is thought to be separate from space-based as well as object-based attention. Indeed, attending to a feature in one location makes it difficult to

ignore the same feature simultaneously presented in another location. For example, Saenz, Buracas, and Boynton (2002) measured fMRI activity in motion- and color-sensitive areas of the human brain and found that attending to a particular direction of motion or a particular color in one location modulated activity for the same features presented in distant locations. In the motion study, they presented a set of dots with half the dots moving upward and half moving downward, and directed attention to either the upward or downward movement during a block of trials. The speed of the attended dots varied slightly from trial to trial. The interwoven dots were presented in one visual field throughout a block of trials, and participants were told that they should pay attention to that side and to ignore the other side. The ignored side contained a set of dots that all moved either upward or downward. Thus they were either moving in the same direction as the attended motion or in the opposite direction. When the movement in the ignored field was the same as the selected movement in the attended field, activity in motion-sensitive areas in the hemisphere contralateral to the ignored side was greater than when the motion was different. The same pattern occurred in the color condition in color-sensitive areas. An attended feature produced more activity in areas specialized for that feature whether spatial attention was oriented to the stimulus location or not (also see Hopf, Schoenfeld, & Heinze, 2005). In other words, the feature was attended whether it was on the attended or ignored side. Spatial attention did not interact with feature attention.

If spatial and feature attention are separate, feature-based attention should be present in patients with complete spatial deficits, and this is the case. For instance, a red circle presented among blue and yellow circles of about the same luminance attracted the attention of the Balint's patient RM. Although features were not detected as rapidly or as accurately as for normal observers, the number of distractors in the display did not affect performance (L. Robertson, Treisman, Friedman-Hill, & Grabowecky, 1997). In other words, features did not "pop out" as rapidly as for normal observers, but they did continue to pop out. Importantly, when we asked RM where the feature was located, he was at chance, and he was reluctant to make a judgment, saying things like "You know I don't know where it is." This response occurred even when it remained on the screen until he responded.

We also wondered whether he saw only one feature at a time, similar to his simultanagnosia for objects, but this did not seem to be the case. Marcia Grabowecky and I showed RM a series of displays containing 9 colored circles arranged in a 3×3 matrix. He was shown 180 of these matrices, 60 containing one red or one green circle among 8 blue circles, 60 containing 2 red or 2 green circles among 7 blue ones, and 60 containing one red and one green circle among 7 blue ones. We asked RM to report how many circles were green and/or red. When there was only 1 red or 1 green circle, he correctly reported there was one 90% of the time (7% omissions). When there were 2 colored circles, with 1 green and 1 red, he correctly reported there were two 83% of the time (7% omissions). However, the most interesting finding was when the 2 circles were both green or both red; his accuracy rate dropped to 10%, reporting he saw only one (3% omissions). Afterward we asked him to simply report how many circles he saw in each display and what the colors were. He reported all the colors that were present on each trial but said he saw only 1 circle of all three colors. For example, when shown a display with 1 red among 8 blue, he reported seeing 1 red and 1 blue circle, which was the same response he gave when shown a display with 2 red circles among 7 blue. When shown a display with 1 red, 1 blue, and 7 green, he easily named the 3 colors but said there was only one of each. Subsequently, when asked if there was more blue on the page, he said no.

The undamaged portions of RM's brain encoded basic features in the visual scene but not their location or their relative magnitude. The result was a chaotic perceptual world in which objects appeared and disappeared without warning, and features coded by specialized neural populations were not bound to a location. The result was a high rate of illusory conjunctions (see Treisman, chapter 12). In the absence of space, binding the wrong color, size, or motion with the object of attention is prominent (Bernstein & Robertson, 1998; Friedman-Hill, Robertson, & Treisman, 1995; Humphreys, Cinel, Wolfe, Olson, & Klempen, 2000; L. Robertson et al., 1997).

Consistently, visually searching for a target that was the conjunction of two features was severely compromised (presumably due to RM's binding problem). He was very poor at reporting the presence or absence of a red *X* among only 3 distractors (e.g., red *O* and 2 green *X*'s) and was extremely frustrated by the task. More tellingly, when the red *X* was absent from the display (e.g., two red *O*'s and 2 green *X*'s), he reported that the target was present 38% of the time (L. Robertson et al., 1997). This high false alarm rate was especially revealing when compared to his 4% miss rate when the target was present. When asked if he was reporting what he actually saw, assured us that he was.

RM's description of his visual experiences in everyday life was consistent with a binding problem as well. For example, he reported he saw a house on his street moving sideways. It is quite possible that the motion from a passing car was detected and then bound incorrectly with the object of his attention (the house).

Implicit Spatial Maps Any task that engages spatial attention obviously will be disrupted when the spatial map

that guides it is damaged. Other studies with Balint's patients have corroborated this conclusion. For instance, utilizing a cue to attend to a given location in anticipation of an upcoming target (Posner, 1980) is all but impossible (L. Robertson & Rafal, 2000), and shifting attention from local to global levels of a display is disrupted as well.

Nevertheless, exogenous spatial orienting appears to be intact, and priming measures have shown that shape at a global level of a hierarchically structured pattern is implicitly represented even when it is not perceived (Egly, Robertson, Rafal, & Grabowecky, 1995; Karnath, Ferber, Rorden, & Driver, 2000). In addition, priming studies have shown that implicit spatial information is present in Balint's patients (Kim & Robertson, 2001; L. Robertson et al., 1997). For example RM read the word "up" faster when it was in the upper part of a rectangle than the word "down" in the same location and vice versa, while he was at chance in reporting the word's location. This result leads to the question of why these spatial maps are not accessible after damage to bilateral occipital-parietal areas.

There is ample evidence from neurobiological studies for the existence of multiple spatial maps throughout the visual cortex. Topographical tuning of visual neural responses in animals and fMRI BOLD responses in humans have been mapped from the fine spatial resolution of V1 to the rough spatial resolution of the parietal lobes, as well as in posterior temporal lobes and the frontal eye fields (Anderson, Batista, Snyder, Buneo, & Cohen, 2000; Colby & Goldberg, 1999; Desimone & Duncan, 1995; Graziano & Gross, 1994; Grill-Spector & Malach, 2004; Laeng, Brennen, & Espeseth, 2002; Silver, Ress, & Heeger, 2005; Wandell, Brewer, & Dougherty, 2005). Gross and Graziano (1995) argued that the parietal lobe functions as a selection hub and noted its strong connections to several other areas of the brain that contain topographical maps. Basically, the suggestion is that the parietal lobe is the gatekeeper in a network of spatial maps that decides which spatial map to access for the task at hand. The loss of parietal functioning would deny access to remaining spatial maps through disconnection.

Another possibility is that parietal functions integrate spatial information from other areas that contain topographical information, which then emerges into what Treisman (1988) called a "master map of locations," and it is this map that guides voluntary spatial attention (L. Robertson, 2003). It is also this map that allows for the perception of a unified spatial world. In this scenario, the loss of parietal function directly damages the master map and consequently the necessary spatial information for object individuation, feature colocation, perceptual organization, and of course the voluntary control of spatial attention. The hypothesis is that spatial information that feeds into the computation of the master map stays below the level of spatial awareness even in normal perception to attenuate the possibility of spatial confusion. Further research is needed to sort out which of these accounts is more plausible.

Consequences of partial spatial loss and attention

Partial spatial loss is much more frequent than Balint's syndrome, and as a result more is known about how syndromes such as unilateral neglect affect perception and attention. In addition to the many scientific papers on the subject, there are several books and chapters that discuss the symptoms, diagnosis, rehabilitation, and/or natural course of recovery over time (see DeRenzi, 1982; Driver, Veuilleumier, & Husain, 2004; Karnath, Milner, & Valler, 2002; Heilman, Watson, & Valenstein, 2003; I. Robertson & Halligan, 1999). Although the neuroanatomical damage that produces unilateral visual neglect is thought to be more anterior than that found in Balint's syndrome (see L. Robertson, 2004), the two spatial problems can produce similar spatial attention deficits that can affect perception in similar ways, but obviously more severely on the contralesional than ipsilesional side for neglect.

Studies of patients with unilateral visual neglect have shown that feature search, although longer on the neglected than unneglected side, remains relatively intact. The difference in search rate for features on the contralesional and ipsilesional sides for patients with neglect is not a result of the fact that a normally parallel search turns into a serial search, since adding distractors on the neglected side does not change response time or accuracy of feature detection (Brooks, Wong, & Robertson, 2005; Esterman, McGlinchey-Berroth, & Milberg, 2000). Conversely, searching for the conjunction of two features (requiring binding and endogenous control of attention) is either not initiated on the neglected side or substantially delayed (often taking a minute or more to begin). Most importantly, the time required to make a decision about the presence or absence of a target on the contralesional side increases as the number of distractors increases (Eglin, Robertson, & Knight, 1989; Esterman, et al.; Laeng et al., 2002; Pavolovskya, Ring, Groswasser, & Hochstein, 2002; Riddoch & Humphreys, 1987b). These findings demonstrate that not all information on the neglected side fails to attract attention. Features that are coded independent of spatial awareness continue to pop out, while conjunctions that require controlled spatial attention do not.

Illusory conjunctions are also more likely on the neglected side of space in cases of unilateral neglect (Cohen & Rafal, 1991), similar to Balint's syndrome, but of course limited to the contralesional side. The most apparent problem in both syndromes is a deficit in spatial attention, and this fact led to the idea that Balint's syndrome was the bilateral version of unilateral neglect. However, this seems not to be the case, as there are important differences between the two

FIGURE 18.3 Example of object-based neglect in drawing of a chair next to a fireplace.

syndromes other than their bilateral/unilateral incarnation. For instance, between 50% and 80% of patients with unilateral neglect also show evidence of object-based neglect (figure 18.3), while patients with Balint's syndrome see nothing but objects, even if only one at a time. Their commonality in disrupting spatial attention does not affect object-based attention in a common way. Also, neglect often disrupts the very concept of space on the neglected side. Patients with neglect act as if that side does not exist, while patients with Balint's syndrome know there is a space out there; they just cannot see it. This observation leads to the possibility that one syndrome represents direct damage to spatial selection, whereas the other represents the loss of a master spatial map on which selection relies.

IMPLICIT PROCESSING AND NEGLECT There has been much interest in what gets processed outside the focus of attention, both in normal observers and in patients with attentional disorders. The case of unilateral neglect presents a situation in which the effects of spatial awareness can be studied in the same individual by comparing performance when stimuli are presented on the ipsilesional and contralesional sides of the visual field. It also provides an opportunity to explore the level at which perceptual processing takes place when the very existence of a sensory event in neglected space is absent.

As with studies of Balint's syndrome, priming methods have shown that a great deal of visual information exists below the level of awareness in the neglected field. For instance, an undetected line drawing of a baseball bat presented on the neglected side speeds the ability to determine whether the letter string "baseball" is a word or not (McGlinchey-Berroth, Milbert, Verfaellie, Alexander, & Kilduff, 1993). The semantics of the undetected object (the baseball bat) primes the word decision response.

Priming from undetected objects in the neglected field can be observed even after almost an hour delay between the prime and probe. Vuilleumier, Schwartz, Clarke, Husain, and Driver (2002) showed primes on the neglected side that were complete line drawings of objects followed 50 minutes later by probes that were either new or from the prime list. The probes varied in the number of line segments used to draw each object (from sparse to dense) and were presented where the stimuli were clearly visible. Fragmentation threshold was measured and defined as the number of fragments that were necessary for the patients to identify the probe. Thresholds for primes that were not detected at all were lower for old than for new objects and closer to primes that had been seen during the prime phase (although still significantly higher).

Findings such as these show that undetected objects are processed up to and including semantic knowledge and suggest that visual objects are bound as a whole before attention is engaged. They have been used to argue that attention simply acts to modulate information, boosting these preattentively bound items above some threshold for awareness. However, as the next section will show, this is not always the case. Attention does have a role to play in addition to modulation.

IMPLICIT PROCESSING, BINDING, AND ATTENTION The evidence for implicit representations of objects as a whole in the neglected field is consistent with results of studies with normal observers. For instance, using psychophysical measures, Breitmeyer, Ogmen, Ramon, and Chen (2005) varied the time between a shape and mask and examined priming effects for wholes and parts. In one case the shapes were "invisible," and in the other case they were "visible." The prime shapes were a square and diamond with one being shown on each trial followed by a mask that was either congruent or incongruent in shape with the prime (figure 18.4). The primes were either wholes (had connected contours) or parts (e.g., corners). The primes were shown for 13 ms followed by a mask 40 or 200 ms latter. At 40 ms participants were unable to identify the primes any better than chance, but at 200 ms, accuracy in discriminating the primes was about 95%. This difference in accuracy was

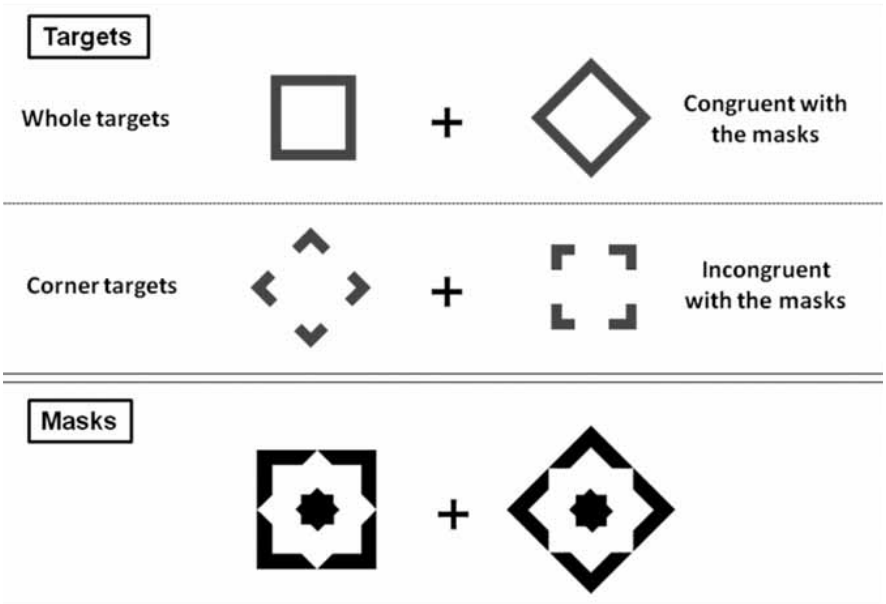


FIGURE 18.4 Example of stimuli used by Breitmeyer, Ogmen, Ramon, and Chen (2005) to study preattentive binding of parts into shapes.

about the same whether the primes were wholes or parts. But the most interesting finding was that congruency between the prime and mask influenced reaction times to report whether the mask was a square or a diamond even in the 40-ms condition where the primes were invisible. Whether the primes were wholes or corners, responses to the mask shape in the congruent conditions were faster than those in the incongruent conditions. The results are consistent with an implicit holistic representation of the prime. However, there was a difference between priming by wholes and parts that depended on visibility, with the congruency effects being larger for wholes than parts when the primes were invisible, but the reverse when they were visible. At the preattentive level, shape representations were stronger when the primes were wholes than when they were parts, while the reverse was true when they were visible. These results suggest that more processing was needed in perceptually binding the corners into a square when the primes were visible, possibly because of their weaker representation at the preattentive level.

A more recent study by Tapia, Breitmeyer, and Schooner (in press) used a similar design to investigate effects of target/mask congruency in binding color and shape and found no evidence for preattentive binding in this case. The primes were again square or diamond shapes, but now all had connecting contours, and the "parts" were shape and color. The shapes were either blue or green on each trial, and the mask was congruent with either the color, the shape, both, or neither. When participants were instructed to respond to the color only or shape only, prime visibility did not affect the degree of priming. However, when they responded to the combination of feature and shape (i.e., conjunctions), there was no evidence to support conjunction priming in the invisible condition, whereas priming was evident in the visible condition.

These results are also consistent with other behavioral findings reported by Lavie (1997) examining the effects of focused and distributed attention on feature and conjunction processing using a flanker task. Three colored shapes were arranged horizontally across the screen with a target appearing either in the same location on every trial (focused attention) or in one of the three locations (divided attention). When participants were focused on a location throughout the trial, congruency between the target and flanker features (color or shape) influenced response time, with incongruent features causing more interference than congruent features. However, flankers that contained the combination of the two target features were no more influential than the features alone. Conversely, when attention was divided across the display, both incongruent feature and incongruent conjunction flankers interfered with response time. Together these studies suggest that shapes may be bound preattentively, but that features that are properties of a shape (see Treisman, 1996) such as color require spatial attention to be correctly bound.

Tom Van Vleet and I recently pursued this issue by studying three patients with left neglect resulting from right hemisphere stroke in the middle cerebral artery distribution. In a priming study we presented feature or conjunction displays as primes in the periphery followed by probes presented at fixation. Using a staircase procedure, we first determined how long a search display had to be presented (threshold

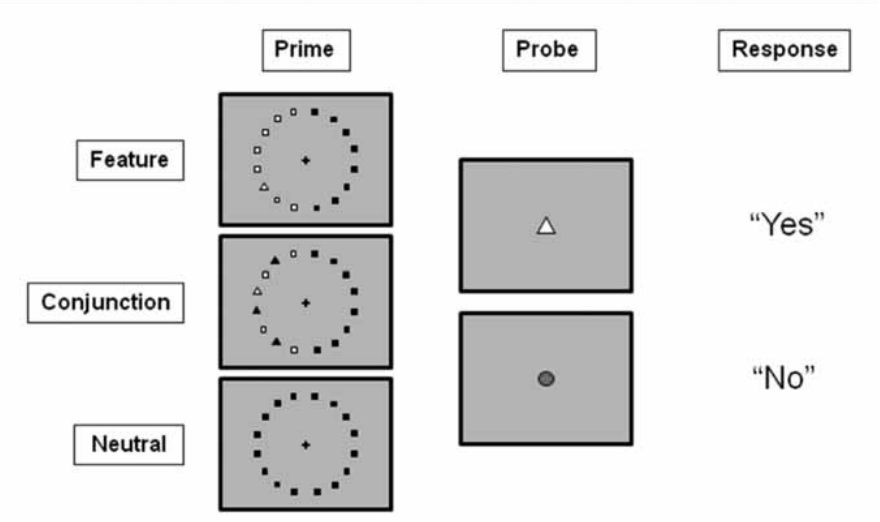


FIGURE 18.5 Example of stimuli used by VanVleet and Robertson in a study of patients with unilateral left neglect/extinction. Dark items in the figures were actually "blue," white items were actually "red," and the gray circle probe was actually "yellow."

presentation time, TPT) to produce a high (75%) or low (25%) probability of target detection for each patient for features on the one hand and conjunctions on the other (figure 18.5). In this way, we equated for detection of features and conjunctions at two levels of difficulty: one in which targets were detected most of the time and another in which they were missed most of the time. We then used the resulting TPTs in a priming stage of the experiment, showing the primes (the same feature and conjunction displays) at the estimated high and low detection TPTs, randomly presented throughout a block of trials. The prime was followed 500 ms later by a single colored shape in the center of the screen, and the patient was asked to respond yes or no as rapidly as possible whether it was a red triangle or not. Priming effects were calculated by subtracting responses to the red triangle when the prime was neutral from when it was either a feature or conjunction target. Priming effects were greater for the conjunction condition when the target was more often visible (high TPT) than when it was more often invisible (low TPT). Conversely, although there was significant priming in the feature condition, it did not differ for high and low TPT (figure 18.6). Follow-up studies demonstrated that neither the differences in the number of red and blue items in the feature and conjunction displays shown in figure 18.5 nor the color differences between the right and left sides of the display could account for these effects.

The evidence from normal observers and patients with unilateral neglect converges to support the conclusion that features coded by specialized neural populations are integrated through spatial attentional control and that parietal functions are critical. Features themselves prime a subsequent response whether or not they are likely to be detected. The studies in normal observers also suggest that there is a difference between preattentive binding of features such as color and shape and preattentive binding of parts into shape. Top-down control is involved in binding depending on what is to be bound. The visual system appears to work independent of spatial awareness when integrating parts into whole shapes but seems to require spatial attention to bind across separated feature maps. It has been suggested that the neural signal involved in binding produces increased gamma band responses in the EEG (Womelsdorf & Fries, chapter 20, this volume). Consistently, Landau, Esterman, Robertson, Bentin, and Prinzmetal (2007) showed that voluntary spatial attention produced greater induced gamma than did automatic spatial orienting. Likewise, increased BOLD activity in the fusiform face area has been observed when a cue predicts the location of an upcoming target face but not when the same cue is unpredictive (Esterman et al., 2008). Voluntary attention seems to increase the perceptual fidelity of the target, while involuntary does not (Prinzmetal, McCool, & Park, 2005).

FEATURE INTEGRATION THEORY (FIT) REVISITED According to feature integration theory as it was originally proposed by Treisman and Gelade (1980), the reason spatial attention was engaged in conjunction search but not feature search was the binding requirements in detecting a conjunction target among multiple items with similar features. Some investigators have argued that it is not the binding requirements per se but rather differences in difficulty (some call it saliency) between targets embedded in the two types of search displays. When difficulty is high, more top-down mechanisms that control attention must be

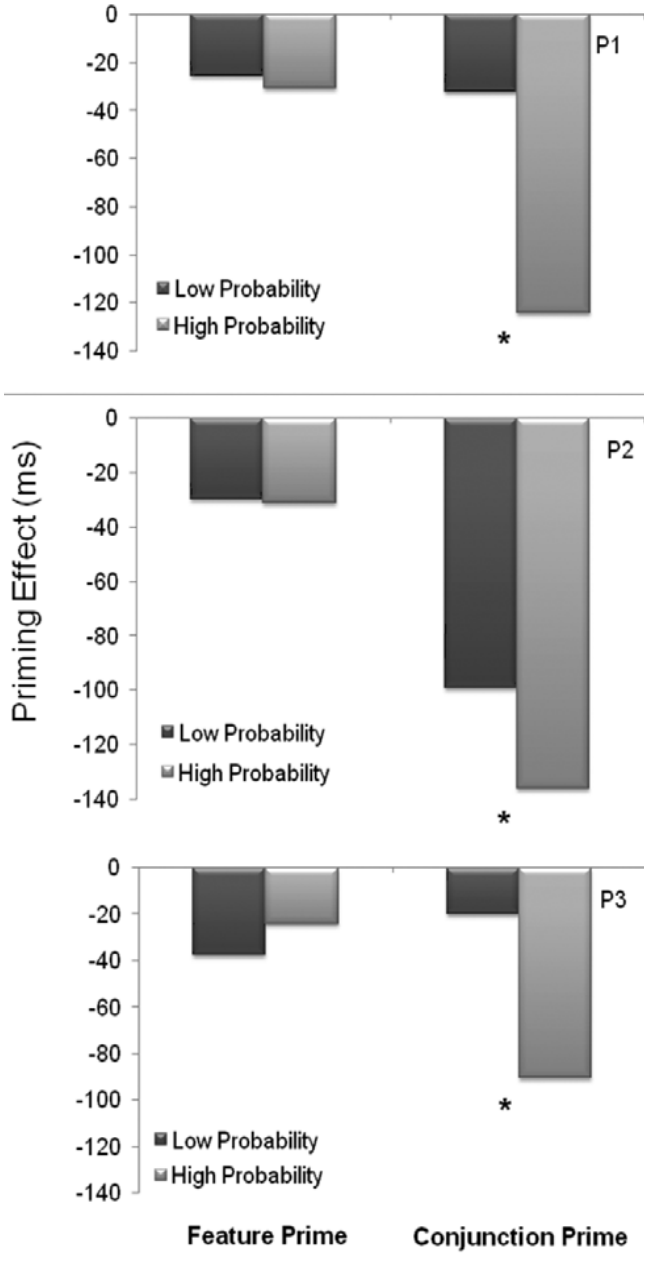


FIGURE 18.6 Differences in reaction time to respond to the central target as a function of prime type (respective search prime minus neutral prime conditions) for three different patients.

engaged than when difficulty is low, and when more top-down attentional control is needed, parietal activity will be higher. There is general agreement that dorsal frontal-parietal attentional systems are activated under most conditions when top-down attentional control is needed to perform a task.

The related question is whether the typical differences in difficulty between feature and conjunction search are a sufficient explanation or whether something else is required, namely, a binding process. The study of neglect described earlier equated search difficulty at a high and low detection threshold, yet found differences in priming between feature and conjunction displays. There is also fMRI evidence that binding can be separated from search difficulty. Donner and colleagues (2002) presented normal participants with conjunction and feature search displays, with the feature search being either the same as or different from the conjunction search in difficulty. Consistent with a difficulty component for search, they found more parietal and frontal activity for both the hard-feature and conjunction-search tasks than for the easy-search task. Nonetheless, they also found an area of activation that could not be explained by difficulty. Activity at the junction of the posterior inferior parietal sulcus and dorsal occipital lobe was more pronounced with conjunction search than with hard-feature search even though behaviorally they were equally difficult. These findings are consistent with a feature-binding mechanism that is engaged in conjunction search but not feature search. They also support FIT in that parietal activity increases whenever a serial search is initiated, but parietal functions are also involved when feature integration is required. The findings are also consistent with previous results using PET to examine the neurobiology of conjunction and feature search. For instance, Corbetta, Shulman, Miezin, and Petersen (1995) showed search displays and varied the task to look for either a feature (motion or color) or the conjunction of motion and color. They found increased activity in both posterior temporal and superior parietal lobes when participants responded to the conjunction of motion and color but only temporal activation when they responded to only motion or only color.

Conclusions

Even when the representation of external space disappears completely with damage to both parietal lobes, anatomically intact areas outside these regions continue to respond to objects and object parts as well as basic features encoded by specialized neural populations (e.g., color, size, motion). However, the locations of these objects and their surface features may not be accurately bound together in perception. Studies of the interaction between spatial maps, attention, and binding with patients who suffer spatial loss (both complete and partial) have contributed to a better understanding of the neural systems involved in the perception of a normally unified spatial world. They also support a role for this spatial map in guiding spatial attention and in perceiving properly bound features. These studies have also clarified what a patient with spatial loss may and may not

see as well as what representations below the level of awareness remain intact.

REFERENCES

Albrecht, A., List, A., & Robertson, L. C. (2008). Attentional selection in the representation of holes and objects. *J. Vis.*, *8*(13), 1–10.

Anderson, R. A., Batista, A. P., Snyder, L. H., Buneo, C. A., & Cohen, Y. E. (2000). Programming to look and reach in posterior parietal cortex. In M. Gazzaniga (Ed.), *The new cognitive neurosciences*. Cambridge, MA: MIT Press.

Balint, R. (1909). Seelenlähmung des "Schauens," optische Ataxie, räumliche Störung der Aufmerksamkeit. *Monatsschr. Psychiatr. Neurol.*, *25*, 5–81. (Translated in *Cogn. Neuropsychol.*, *12*, 265–281, 1995.)

Bartolomeo, P., & Chokron, S. (2001). Levels of impairment in unilateral neglect. In M. Behrmann (Ed.), *Disorders of visual behavior*. (Vol. 4, pp. 67–98). Amsterdam: Elsevier Press.

Bartolomeo, P., Thiebaut de Schotten, M., & Doricchi, F. (2007). Left unilateral neglect as a disconnection syndrome. *Cereb. Cortex*, *17*, 2479–2490.

Bernstein, L. J., & Robertson, L. C. (1998). Independence between illusory conjunctions of color and motion with shape following bilateral parietal lesions. *Psychol. Sci.*, *9*, 167–175.

Breitmeyer, B. G., Ogmen, H., Ramon, J., & Chen, J. (2005). Unconscious and conscious priming by forms and their parts. *Visual Cogn.*, *12*, 720–736.

Brooks, J. L., Wong, Y., & Robertson, L. C. (2005). Crossing the midline: Reducing visual extinction by re-establishing hemispheric balance. *Neuropsychologia*, *43*, 572–582.

Cohen, A., & Rafal, R. (1991). Attention and feature integration: Illusory conjunctions in a patient with parietal lesion. *Psychol. Sci.*, *2*, 106–110.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annu. Rev. Neurosci.*, *22*, 319–349.

Corbetta, M., Shulman, G., Miezin, F., & Petersen, S. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunctions. *Science*, *270*, 802–805.

DeRenzi, E. (1982). *Disorders of space exploration and cognition*. New York: Wiley.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, *18*, 193–222.

Donner, T. H., Kettermann, A., Diesch, E., Ostendorf, F., ffillinger, A., & Brandt, S. (2002). Visual feature and conjunction search of equal difficulty engage only partially overlapping frontoparietal networks. *NeuroImage*, *15*, 16–25.

Driver, J., Veuilleumier, P., & Husain, M. (2004). Spatial neglect and extinction. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (3rd ed.). Cambridge, MA: MIT Press.

Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.*, *113*, 501–517.

Eglin, M., Robertson, L. C., & Knight, R. T. (1989). Visual search performance in the neglect syndrome. *J. Cogn. Neurosci.*, *4*, 372–381.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence for normal and parietal-lesion subjects. *J. Exp. Psychol. Gen.*, *123*, 161–172.

Egly, R., Robertson, L. C., Rafal, R., & Grabowecky, M. (1995). *Implicit processing of unreportable objects in Balint's syndrome.* Presented at Psychonomic Society meeting, Los Angeles.

Esterman, M., McGlinchey-Berroth, R., & Milberg, W. P. (2000). Parallel and serial search in hemispatial neglect: Evidence for preserved preattentive but impaired attentive processing. *Neuropsychology*, *14*, 599–611.

Esterman, M., Prinzmetal, W., DeGutis, J., Landau, A., Hazeltine, E., Verstynen, T., & Robertson, L. (2008). Voluntary and involuntary attention affect face discrimination differently. *Neuropsychologia*, *46*, 1032–1040.

Friedman-Hill, S., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science*, *269*, 853–855.

Graziano, M. S. A., & Gross, C. G. (1994). Mapping space with neurons. *Curr. Dir. Psychol. Sci.*, *3*, 164–167.

Grill-ˢpector, K., & Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.*, *27*, 649–677.

Gross, C. G., & Graziano, M. S. A. (1995). Multiple representations of space in the brain. *Neuroscientist*, *1*, 43–50.

Heilman, K. M., Watson, R. T., & Valenstein, E. (1994). Localization of lesions in neglect and related disorders. In A. Kertesz (Ed.), *Localization and neuroimaging in neuropsychology*. San Diego: Academic Press.

Heilman, K. M., Watson, R. T., & Valenstein, E. (2003). Neglect and related disorders. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (4th ed.). New York: Oxford University Press.

Hof, P. R., Bouras, C., Constantinidis, J., & Morrison, J. H. (1989). Balint's syndrome in Alzheimer's disease: Specific disruption of the occipito-parietal visual pathway. *Brain Res.*, *493*, 368–375.

Holmes, G., & Horax, G. (1919). Disturbances of spatial orientation and visual attention with loss of stereoscopic vision. *Arch. Neurol. Psychiatry*, *1*, 385–407.

Hopf, J.-M., Schoenfeld, M. A., & Heinze, H.-J. (2005). The temporal flexibility of attentional selection in the visual cortex. *Curr. Opin. Neurobiol.*, *15*, 183–187.

Humphreys, G. W., Cinel, C., Wolfe, J., Olson, A., & Klempen, N. (2000). Fractionating the binding process: Neuropsychological evidence distinguishing binding of form from binding of surface features. *Vis. Res.*, *40*, 1569–1596.

Ivry, R. B., & Robertson, L. C. (1998). *The two sides of perception.* Cambridge, MA: MIT Press.

Karnath, H.-O., Ferber, S., & Himmelback, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*, *411*, 950–953.

Karnath, H.-O., Ferber, S., Rorden, C., & Driver, J. (2000). The fate of global information in dorsal simultanagnosia. *Neurocase*, *6*, 295–306.

Karnath, H.-O., Milner, D., & Valler, G. (2002). *The cognitive and neural bases of spatial neglect.* New York: Oxford University Press.

Kim, M. S., & Robertson, L. C. (2001). Implicit representations of visual space after bilateral parietal damage. *J. Cogn. Neurosci.*, *13*, 1080–1087.

Laeng, B., Brennen, T., & Espeseth, T. (2002). Fast responses to neglected targets in visual search reflect pre-attentive processes: An exploration of response times in visual neglect. *Neuropsychologia*, *40*, 1622–1636.

Landau, A. N., Esterman, M., Robertson, L. C., Bentin, S., & Prinzmetal, W. (2007). Voluntary and involuntary attention differentially effects EEG activity in the gamma band. *J. Neurosci.*, *27*, 11986–11990.

Lavie, N. (1997). Feature integration and selective attention: Response competition from unattended distractor features. *Percept. Psychophys.*, *59*, 542–556.

Leung, H. C., Gore, J. C., & Goldman-Rakic, P. S. (2002). Sustained mnemonic response in the human middle frontal

gyrus during on-line storage of spatial memoranda. *J. Cogn. Neurosci.*, *14*, 659–671.

McGlinchey-Berroth, R., Milbert, W. P, Verfaellie, M., Alexander, M., & Kilduff, P. T. (1993). Semantic processing in the neglected visual field: Evidence from a lexical decision task. *Cogn. Neuropsychol.*, *10*, 79–108.

Mesulam, M.-M. (1981). A cortical network for directed attention and unilateral neglect. *Ann. Neurol.*, *4*, 309–325.

Mort, D. J., Malhotra, P., Mannan, S. K., Rorden, C., Pambakian, A., Kennard, C., et al. (2003). The anatomy of visual neglect. *Brain*, *126*, 1986–1997.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.

Pavlovskaya, M., Ring, H., Groswasser, Z., & Hochstein, S. (2002). Searching with unilateral neglect. *J. Cogn. Neurol.*, *14*, 745–756.

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.*, *32*, 3–25.

Prinzmetal, S., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *J. Exp. Psychol.*, *134*, 73–92.

Rafal, R. (1997). Balint syndrome. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral neurology and neuropsychology*. New York: McGraw Hill.

Rafal, R. (2001). Balint's syndrome. In M. Behrmann (Ed.), Disorders of visual behavior (Vol. 4, pp. 121–141). Amsterdam: Elsevier Science.

Riddoch, M. J., & Humphreys, G. W. (1987a). A case of integrative visual agnosia. *Brain*, *110*, 1431–1462.

Riddoch, M. J., & Humphreys, G. W. (1987b). Perception and action systems in unilateral visual neglect. In M. Jeannerod (Ed.), *Neuropsychological and neurophysiological aspects of spatial neglect*. Amsterdam: New Holland.

Rizzo, M., & Vecera, S. P. (2002). Psychoanatomical substrates of Balint's syndrome. *J. Neurol. Neurosurg. Psychiatry*, *72*, 162–178.

Robertson, I. H., & Halligan, P. W. (1999). *Spatial neglect: A clinical handbook for diagnosis and treatment.* East Sussex, UK: Psychology Press.

Robertson, L. C. (2003). Binding, spatial attention and perceptual awareness. *Nat. Rev. Neurosci.*, *4*, 93–102.

Robertson, L. C. (2004). *Space, objects, minds and brains*. New York: Psychology Press (Francis & Taylor), Essays in Cognitive Science.

Robertson, L. C., & Rafal, R. (2000). Disorders of visual attention. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press.

Robertson, L. C., Treisman, A., Friedman-Hill, S., & Grabowecky, M. (1997). The interaction of spatial and object pathways: Evidence from Balint's syndrome. *J. Cogn. Neurosci.*, *9*, 295–317.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat. Neurosci.*, *5*, 631–632.

Silver, M. A., Ress, D., & Heeger, D. J. (2005). Topographical maps of visual spatial attention in human parietal cortex. *J. Neurophysiol.*, *94*, 1358–1371.

Tapia, E., Breitmeyer, B., & Schooner, C. (in press). Role of task-directed attention in nonconscious and conscious response priming by form and color. *J. Exp. Psychol. Hum. Percept. Perform.*

Treisman, A. (1988). Features and Objects: The Fourteenth Bartlett Memorial Lecture. *Q. J. Exp. Psychol. [A]*, *40*, 201–237.

Treisman, A. M. (1996). The binding problem. *Curr. Opin. Neurobiol.*, *6*, 171–178.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psych.*, *12*, 97–136.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in perception of objects. *Cogn. Psych.*, *14*, 107–141.

Vuilleumier, P., Schwartz, S., Clarke, K., Husain, M., & Driver, J. (2002). Testing memory for unseen visual stimuli in patients with spatial neglect and extinction. *J. Cogn. Neurosci.*, *14*, 875–886.

Wandell, B. A., Brewer, A. A., & Dougherty, R. F. (2005). Visual field map clusters in human cortex. *Philos. Trans. R. Soc. London B Biol. Sci.*, *360*, 693–707.

19 The Effect of Attention on the Responses of Individual Visual Neurons

JOHN H. R. MAUNSELL

ABSTRACT Performance on sensory tasks depends not only on the quality of the sensory signals that are available, but also on the aspects of the sensory signals that the subject attends to. Recordings from individual neurons in trained, behaving monkeys have shown that attention to particular visual stimuli alters the way that those stimuli are represented in cerebral cortex. The primary effect of attention appears to be a gain change, which increases the responses of neurons that represent attended stimuli while decreasing the responses of other neurons. This gain change affects responses to all stimuli proportionately, without affecting the selectivity of neurons or the stimulus that they prefer. This effect alone can explain much of the improvement in behavioral performance that is conferred by attention.

One of the most useful features of a personal computer is its ability to multitask. A computer can simultaneously input text for a document, check the spelling and grammar of recently entered words and phrases, copy the text to disk, check for updates to its software, put up reminders about appointments, and do dozens of other important chores. However, in its low-level hardware, the computer is a serial device. All its tasks are accomplished by a central processing unit (or a few central processing units) stepping rapidly through a sequence of instructions. For the most part, parallel processing in a computer is an illusion created by a central processing unit doing one task at a time, but switching between dozens of different tasks so rapidly that the user does not notice. Much of the effort in computer design during the last few decades has been directed at making the hardware of computers more genuinely parallel by adding more central processing units and by delegating minor tasks to peripheral devices.

In contrast to a computer, the neurons and synapses that make up the hardware of the brain process information in a massively parallel way. Sensory signals are collected and analyzed simultaneously by hundreds of millions of neurons in the sensory epithelia, and broad fiber tracts convey the results to more central structures that carry out further parallel computations. The brain uses parallel hardware throughout. Even the simplest behavioral response depends on the concerted activity of thousands of motor neurons.

It seems ironic that while computer engineers strive to make the serial hardware at the heart of a computer emulate parallel behaviors, the brain, with a low-level architecture that seems to be the embodiment of parallel processing, completes tasks in a largely serial way. Unlike a multitasking computer, people and animals generally do one thing during any brief interval. While certain vegetative functions can operate autonomously and in parallel, the higher function of the central nervous system seems to be largely limited to one task at a time.

There are physical constraints on what a single organism can do at one time, but no aspect of body design would prevent a person from, say, doing completely independent tasks with each hand. Nevertheless, the brain does most tasks one at a time. This serial nature of the brain is often described in terms of attention. Attention determines which sensory signals control behavior. We attend to one item or group at a time and shift attention from one subject to another to accomplish our goals. Thus attention is a key player in the brain's serial processing of tasks, and it limits the rate at which information can be processed. For example, evidence from split-brain patients suggests that visual search goes faster when each cerebral hemisphere employs an independent focus of attention (Luck, Hillyard, Mangun, & Gazzaniga, 1989).

What is attention, and how does it affect the processing of sensory signals in the brain? Some of the most detailed information about the role of attention in controlling sensory processing has come from studies of the visual system. Here we will focus on findings from studies that examine how attention affects some of the low-level hardware that supports behaviors: the responses of individual units of visual cerebral cortex in monkeys.

JOHN H. R. MAUNSELL Department of Neurobiology, Harvard Medical School, Howard Hughes Medical Institute, Boston, Massachusetts

How severe is the attentional bottleneck?

Before considering how attention affects neuronal signals, it is useful to consider how much attention affects behavioral responses. Attention is often viewed as a bottleneck through which only a small fraction of sensory signals can pass to reach a stage where they are fully processed. Articles describing experiments on attention often note that the brain can only process a tiny fraction of the welter of sensory information that it receives, and they suggest that attention is the mechanism by which the brain selects the most important signals. Such statements evoke the common experience of arriving home with no recollection of the drive from work. However, an inability to produce details of a recent experience may have little to do with attention. Brains do not have the capacity to record every sensory experience. A failure to recollect events is as likely to reflect the limits of memory, recall, or report as it is the limits of attention at the time those events occurred.

When attention is studied with methods designed to isolate it from memory and other factors, the magnitude of its effects varies considerably. In tasks that involve searching for a stimulus that is hidden among similar distractors, attention can impose a severe bottleneck on sensory processing (Verghese & Pelli, 1992). However, most neurophysiological studies of attention do not explore such challenging situations. Instead, many studies of attention to sensory stimuli use designs related to the Posner paradigm (Posner, 1980), where the behavioral consequences of attention are typically moderate changes in thresholds or reaction times. Even within the controlled conditions of the Posner paradigm, attention may have widely varying effects depending on task demands. Attention might be directed equally to two locations or with different weight to each location depending on reward probabilities (e.g., Mangun & Hillyard, 1990) or the complexity of the task being performed at one location (Lavie & Tsal, 1994; Lavie, 1995). Because the relative amount of attention directed to different stimuli is flexible, behavioral performance may be either severely limited by attention or hardly affected, depending on the stimuli, task, instructions, and reward expectations.

Correspondingly, the effects of attention on neuronal signals can depend on the task a subject performs. Single-unit studies of attention in the monkey visual system have shown that changes in the relative amount of attention allocated to different locations are reflected in the strength of neuronal responses to visual stimuli (Spitzer, Desimone, & Moran, 1988; Spitzer & Richmond, 1991; Basso & Wurtz, 1998; Platt & Glimcher, 1999; Ikeda & Hikosaka, 2003; Boudreau, Williford, & Maunsell, 2006).

A recent study has shown that when task difficulty is increased, the amount that spatial attention changes neuronal responses also increases (Boudreau et al., 2006). In this experiment, a monkey's attention was directed either to a stimulus inside the receptive field of a neuron being recorded in area V4 or to a second stimulus in the opposite hemifield. Responses were typically stronger when attention was directed to the receptive field stimulus than when attention was directed to the other stimulus. However, the difference in response depended on whether the animal was doing an easy or difficult version of the task. When the animal shifted its attention between the two locations in order to detect a 90° change in orientation, responses were modulated relatively little (median 7%, figure 19.1*A*). In interleaved trials

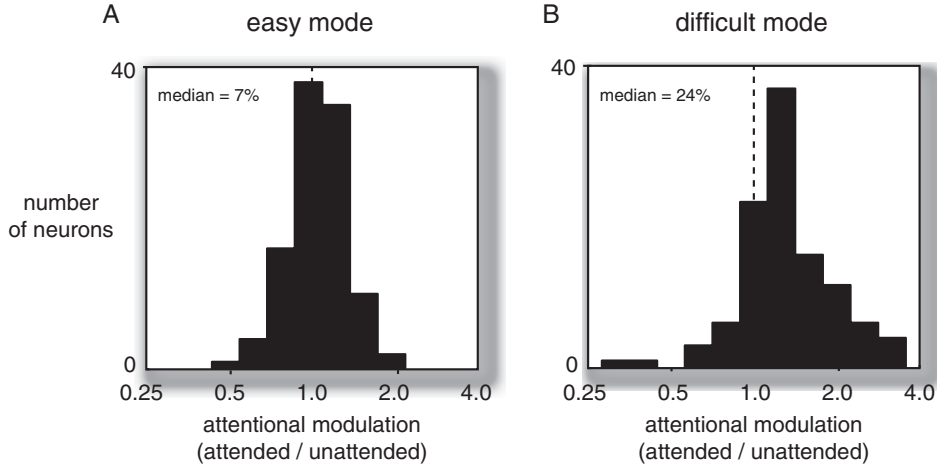


FIGURE 19.1 Attentional modulation of neuronal responses is affected by task difficulty. (*A*) When animals shifted their attention between a stimulus in the receptive field of a V4 neuron and a distant stimulus, the median modulation across all neurons tested was only 7% when the task was easy (detecting an orientation change of 90°). (*B*) When the same neurons were tested during interleaved trials when animals were doing a much more difficult detection (orientation change of ~10°), the median modulation was much greater, 24%. (Data from Boudreau, Williford, & Maunsell, 2006.)

in which the animal was attending to one location or the other to detect a much smaller change in orientation (~10°), the locus of attention affected responses much more (median modulation 24%, figure 19.1*B*).

The amount of attentional modulation of neuronal responses that is observed should be expected to depend on details of the task design. On one hand, because most experiments do not push subjects to the limit of their performance, the magnitude of the attentional modulation of neuronal responses measured in most experiments is almost certainly more modest than it might be. On the other hand, we are rarely pushed to the limits of our abilities in everyday life, so the results reported may be representative of the operating range of attention in typical situations.

How much does attention change the strength of sensory responses?

Because attentional modulation depends on task conditions, there can be no veridical answer to the question of how much attention alters neuronal responses in different parts of visually responsive cerebral cortex. Some general observations are possible, however. First, most studies of attentional modulation of single units describe moderate modulations of rate of firing in all parts of monkey visual cortex, averaging 10–50% (see Maunsell & Cook, 2002). Directing attention toward or away from a stimulus is rarely seen to turn neurons on or off. Functional imaging studies of attentional modulation of neural activity in corresponding regions of human visual cortex have found stronger modulations than those described for single units in monkeys (Kanwisher & Wojciulik, 2000; Pessoa, Kastner, & Ungerleider, 2003). However, EEG and local field potential recordings from human visual cortex typically find moderate effects that are more in keeping with the results from monkey single-unit recording (Hillyard & Anllo-Vento, 1998; Yoshor, Ghose, Bosking, Sun, & Maunsell, 2007), suggesting that the difference may depend more on the indirect measure of neuronal activity used in functional imaging, rather than a species difference. Overall, attention seems to act much more like a moderate filter for sensory representations, rather than a gate, at least in the relatively reduced visual displays that are used in most experiments.

The amount of attentional modulation can depend on stimulus configurations, and some arrangements of visual stimuli consistently produce stronger attentional modulations. Most single-unit studies of attention compare neuronal responses to a given stimulus while attention is directed toward or away from that stimulus. However, shifting attention between two stimuli that are both within a neuron's receptive field can produce stronger modulations. Neurons in later stages of visual cortex generally have receptive fields that are large enough to hold two stimuli that are sufficiently well separated that a subject can direct attention to one or the other. If one stimulus is a preferred stimulus for the neuron and the other is a nonpreferred stimulus, shifting attention between them can produce two- or threefold changes in neuronal response (Moran & Desimone, 1985; Motter, 1994; Luck, Chelazzi, Hillyard, & Desimone, 1997; Treue & Maunsell, 1999).

Whether this stronger modulation reflects a different mechanism than that engaged when a single stimulus is in the receptive field is an important question that remains to be addressed. To date, no experiment has measured attentional modulations with one and two stimuli in the receptive field under comparable conditions. It remains possible that the greater modulation with two stimuli is only apparent, because measurements made with that configuration compare responses with attention to a preferred stimulus with responses with attention to a nonpreferred stimulus. With a single stimulus in the receptive field, the comparison is instead between responses with attention to a preferred stimulus and with responses to a neutral stimulus (one well outside the receptive field that cannot affect responses directly). Additionally, measurements with two stimuli inside a receptive field require closely spaced stimuli, which make the task more difficult. This extra difficulty may contribute to the stronger attentional modulation in this condition. More information about the role of attention with cluttered visual displays will fill an important lacuna in our understanding of attentional modulation. It is possible that the relatively sparse displays used in most experiments produce modest neuronal modulation compared with that occurring in natural viewing conditions. For example, it has been reported that the responses of neurons in inferotemporal cortex are virtually all-or-none depending on whether monkeys notice a target in a cluttered scene (Sheinberg & Logothetis, 2001), but comparisons were not made using equivalent retinal stimulation in the two conditions.

While attention typically has modest effects on neuronal rate of firing in laboratory experiments, attention has other effects on neuronal response. Attention can also modulate the amount of synchrony or gamma power in neuronal activity, both in monkey microelectrode recordings (Steinmetz et al., 2000; Fries, Reynolds, Rorie, & Desimone, 2001; Taylor, Mandon, Freiwald, & Kreiter, 2005; Fries, Womelsdorf, Oostenveld, & Desimone, 2008) and human macroelectrode recordings (see Gruber, Müller, Keil, & Elbert, 1999; Müller, Gruber, & Keil, 2000; Müller & Gruber, 2001; Jensen, Kaiser, & Lachaux, 2007; Wyart & Tallon-Baudry, 2008). Changes in gamma-band activity have been correlated with the behavioral signatures of attention (Womelsdorf, Fries, Mitra, & Desimone, 2006). While these changes in gamma oscillations or synchrony are

modest, such changes could in principle provide a powerful mechanism for attention to affect sensory processing (see Womelsdorf & Fries, chapter 20 in this volume and Tiesinga, Fellous, Salinas, Jose, & Sejnowski, 2004; Boergers, Epstein, & Kopell, 2005; Tiesinga, Fellous, & Sejnowski, 2008).

Does attention affect the selectivity of visual neurons?

Beyond changing the strength of neuronal responses, attention might also improve behavioral performance by changing the selectivity of neurons. All sensory neurons are selective for certain stimuli. Selectivity for continuous dimensions (e.g., orientation, color, stereoscopic depth) is usually described by tuning curves that plot the strength of a neuron's response to different stimulus values. Narrow tuning curves correspond to high selectivity, while broad tuning curves correspond to low selectivity. How the width of tuning in neuronal populations translates to behavioral performance is a poorly understood and complex subject (see Pouget, Deneve, & Ducom, 1999; Butts & Goldman, 2006), but tuning width is likely to be an important factor in determining behavioral performance. For example, the width of orientation tuning curves might be thought of as the "graininess" of the brain's representation of orientation. By narrowing tuning curves, attention might improve discrimination of different stimulus dimensions.

Given the potential advantages of changing neuronal selectivity, it is notable that single-unit studies have found little evidence for such changes. When orientation tuning curves were measured for neurons in area V4 in monkey visual cortex, attention scaled the entire tuning curve to stronger responses without changing the breadth of tuning (Spitzer et al., 1988; McAdams & Maunsell, 1999a). Figure 19.2 shows average orientation tuning curves from a sample of V4 neurons. The solid points are average responses to different orientations of a receptive field stimulus when attention was directed to the stimulus. The open points are responses to the same stimuli when attention was directed elsewhere. Attention scaled the orientation tuning curve vertically by about 30%, but it did not change the width of orientation tuning.

Similar experiments have shown that attention to a moving stimulus does not affect the breadth of direction tuning for neurons in monkey middle temporal area (MT), even when attention is directed to a specific direction (Treue & Martinez-Trujillo, 1999). Recordings of event-related potentials in humans have likewise seen that spatial attention increases the amplitude of response waveforms with little change in their waveform, latency, or distribution across the scalp (Mangun & Hillyard, 1987, 1990).

A proportional scaling of tuning curves by attention is reminiscent of the effects of changing stimulus contrast or intensity, which similarly increases or decreases responses

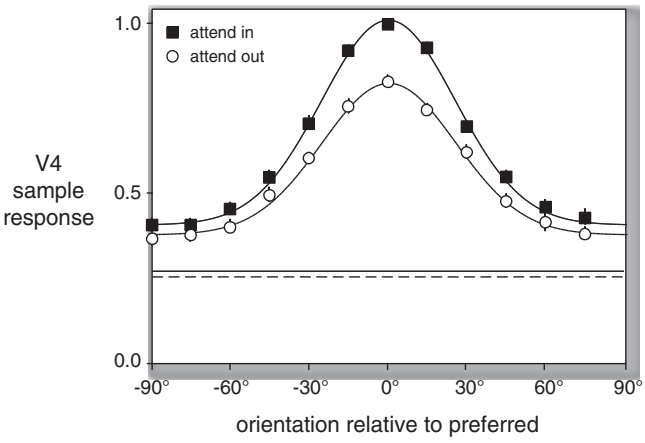


FIGURE 19.2 Average orientation tuning of neurons in area V4 with and without attention directed to the stimulus. The orientation tuning of 262 individual V4 neurons was measured twice: once with the animal attending to the stimulus to report its orientation and once with the animal's attention directed elsewhere. Each neuron's tuning curves were normalized to a maximum response of 1.0 and shifted to align their preferred orientations. Values from both tuning curves were then averaged across cells. Overall, responses were about 30% stronger when the animal was attending to the stimulus in the receptive field (amplitude of the function fitting the response with attention directed to the receptive field, 0.60; with attention directed elsewhere, 0.45). However, width of the tuning curve was unaffected by attention (the width of the fitted functions in both cases was 37°). Horizontal lines mark average spontaneous activity for attended (solid) and unattended conditions (dashed). (Data replotted from McAdams & Maunsell, 1999a.)

without affecting the width of tuning curves (see Troyer, Krukowski, Priebe, & Miller, 1998). This similarity is intriguing because the behavioral advantages conferred by attention, faster responses, and better discrimination are similar to the behavioral consequences of higher stimulus contrast. Thus attention might produce better behavioral performance by effectively increasing the relative intensity of behaviorally relevant stimuli at the level of the cortical representation.

The similarity between the effects of attention and stimulus contrast on cortical responses was reinforced by studies that examined how attention affects neuronal tuning curves for contrast. Most neurons in early stages of visual cortex have sigmoidal tuning curves for contrast. They are insensitive to very low contrasts (lower saturation), produce increasingly strong responses over a middle range of contrasts (rising phase), and then show relatively little change in response as contrast is increased through the highest values (upper saturation). Studies of the effects of attention on contrast response functions in both area V4 (Reynolds, Pasternak, & Desimone, 2000) and MT (Martinez-Trujillo & Treue, 2002) reported that attention had the greatest effect on responses to low contrasts, as if it added a fixed proportion of contrast to stimuli at the attended location

and thereby shifted the contrast tuning curve toward lower values.

However, the effect these reports described was distinct from that described for orientation or direction tuning curves, where attention scaled responses proportionally. If attention scaled contrast tuning curves proportionally, its greatest effect would have been at the highest contrasts, where responses are strongest. A subsequent study of the effect of attention on contrast response functions has called the earlier conclusions into question. Williford and Maunsell (2006) reexamined the effects of attention on contrast tuning curves in V4 and found they could not distinguish whether effects were strongest at low contrast or high contrast. Because most neurons in V4 do not show strong saturation at high contrast, models describing effects primarily at high and low contrast are both able to fit the data well. As noted in that report, while each of the three studies of the effects of attention on contrast tuning curves favored one or the other model, none ruled out the alternative model as an acceptable description.

Thus it remains possible that attention does not have a special effect on stimuli of low contrast, but instead has a single effect on all neuronal tuning curves: a simple proportional scaling of responses to all stimuli, without change in the breadth of tuning or the preferred stimulus. It should be possible to resolve this question by examining neurons with strongly saturating tuning curves and collecting sufficiently precise data to see whether attention has proportionally greater effects at low contrasts.

How does attentional modulation of neuronal responses improve behavior?

What is accomplished by modulating the strength of neuronal responses? One obvious suggestion is that stronger responses can have a better signal-to-noise ratio. Sensory neurons give variable responses. They produce different numbers of spikes in response to different presentations of the same stimulus. For most neurons throughout visual cerebral cortex, the variance in the number of spikes approximates the mean number of spikes for the response, as is true for a Poisson process (Softky & Koch, 1993; Shadlen & Newsome, 1998). If the signal-to-noise ratio of a response is defined as the mean response (signal) divided by the standard deviation of the response (noise), then signal-to-noise is expected to improve systematically as responses are made stronger, because the standard deviation is the square root of the variance, and it therefore remains proportional to the square root of the mean response.

In a study of the effects of attention on the responses to different orientations in V4 (McAdams & Maunsell, 1999b), neurons were found to respond an average of 30% more strongly when attention was directed toward the stimulus in their receptive field. Calculations showed that this increase in response would improve the median neuron's smallest discriminable orientation (for a peripheral stimulus) from 26.5° to 20.4°, by virtue of improved signal-to-noise.

This study also considered whether attention might have a more dramatic effect on signal-to-noise by reducing the variance directly, but found no evidence for an effect on variance beyond that expected from changing the strength of the response. Recently, however, Mitchell, Sundberg, & Reynolds (2007) reexamined the effects of attention in V4 by classifying neurons as broad-spiking (putative pyramidal cells) and narrow-spiking (putative inhibitory interneurons) based on the duration of their action potentials. Although attention seemed to have proportional effects on the rate of firing of these two cell types, there was a difference in its effects on the variance of their responses. Attention did not affect the relationship between response rate and response variance for the broad-spiking neurons, but for the narrow-spiking neurons there was additional variance when a stimulus was unattended. The effect was small and appeared only at high rates of firing, and the responses of narrow-spiking neurons to attended stimuli had no less variance than broad-spiking neurons showed for attended or unattended stimuli. It is not clear what benefits arise from increasing the variance of responses of one class of cells to unattended stimuli.

This unexpected effect of attention on the variance of the responses of narrow-spiking neurons leads to a more general question about attention and neuronal signal-to-noise. If attention to a stimulus can produce better signal-to-noise through either stronger responses or less variance, why would the brain ever want to decrease the signal-to-noise of neuronal responses? One explanation is the cost of high rates of firing. Metabolic expense is a considerable factor for the brain, which consumes an inordinate amount of the body's energy (Attwell & Laughlin, 2001; Lennie, 2003). It may not be practical to maintain the higher rates of firing to achieve higher signal-to-noise. Alternatively, the answer may instead depend on how neuronal signals translate into target detection and false alarms. Most of the higher signal-to-noise that attention achieves is gained through making neurons more sensitive. When the response of a neuron is enhanced by attention, it will give the same response to some nonpreferred stimuli that it would give to a preferred stimulus when its response is not enhanced. Enhanced responses to nonpreferred stimuli could be interpreted as false alarms about the presence of a preferred stimulus. When you search for red stimuli, it may be acceptable, or even adaptive, to have false alarms from red-preferring neurons, but false alarms from red-preferring neurons are more likely to be maladaptive when searching for other colors.

It has been suggested that attention increases the sensitivity of neurons according to how closely their response

properties match the current target of attention (Treue & Martinez-Trujillo, 1999). For example, if a subject were hunting for red targets in a particular portion of the visual field, the responses of neurons with receptive fields overlapping the attended location would be enhanced regardless of their stimulus preference, and the responses of neurons that preferred red would be enhanced regardless of the location of their receptive fields. Neurons that both preferred red and had a receptive field at the attended location would receive the benefits of both effects and would have responses that were more strongly enhanced than any other neurons.

While this appealing suggestion is consistent with physiological data (see Maunsell & Treue, 2006), it does not address the question of why attention modulates neuronal responses by the amount that it does: what is the right amount of attentional modulation? It is obvious that extreme modulation would be unhelpful. If attention has no effects on neuronal activity, then there could be no behavioral effect of attention. Conversely, if attention acted like a gate that silenced neurons representing unattended attributes while activating neurons that represented attended stimuli, it might produce the equivalent of hallucinations. It is safe to assume that the observed modulations have been optimized, but optimized for what?

An insightful answer to this question was recently provided by Navalpakkam and Itti (2007), who noted that enhancing the responses of neurons with properties that match the target of attention may not be optimal if the visual field contains many distracters that are similar to the target. In a computational study, they showed that in many cases optimal performance is achieved by modulating the sensitivity of neurons not based on how well they respond to the target, but based instead on how well they differentiate the target from distracters. In many situations, the neurons that best differentiate target from distracter will prefer neither. Consistent with this finding, in psychophysical studies with human subjects performing a search for lines of a particular orientation, these investigators found evidence for greatest improvement in sensitivity for an orientation that was neither the target orientation nor the distracter orientation. Neurophysiological recording from single units should be able to provide a direct test of this proposal. Because it provides a specific prediction for the amount of modulation that each neuron should receive, it may help explain the substantial differences in neuronal modulation seen within and between cortical areas (see Maunsell & Cook, 2002).

Concluding comments

There are many important questions about the neuronal underpinnings of attention that remain unanswered. Some of these are straightforward and can be answered using existing approaches. For example, we still lack information on the specificity of attentional modulation in different areas of visual cortex. Does attention modulate the activity of all neurons with receptive fields overlapping an attended location? Or, if attention is directed to a location to recognize a face, are neurons in areas specialized for motion processing excluded from its effects? Other questions will require approaches that have been used in studies of attention only infrequently. These include exploring whether attention depends on particular neurotransmitters or neuronal circuits. It will be important to know whether attention is a specialized neural system, or if instead it represents just another type of signal that is intermixed with the bottom-up inputs that make up the core of sensory information processing. Finally, there are conceptual challenges to understanding attention and its role in sensory processing. It has been noted that the type of attention we have considered here is formally indistinguishable from what has been called reward expectation in other lines of experiments that are largely separate from studies of attention (see Sparks, 1999; Roesch & Olson, 2003; Maunsell, 2004). Clarifying definitions, terminology, and concepts may play as important a role in understanding attention as additional experimental data.

ACKNOWLEDGMENTS This work was supported by the National Institutes of Health (R01 EY05911). JHRM is an investigator with the Howard Hughes Medical Institute.

REFERENCES

Attwell, D., & Laughlin, S. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.*, *21*, 1133–1145.

Basso, M. A., & Wurtz, R. H. (1998). Modulation of neuronal activity in superior colliculus by changes in target probability. *J. Neurosci.*, *18*, 7518–7534.

Boergers, C., Epstein, S., & Kopell, N. (2005). Background gamma rhythmicity and attention in cortical local circuits: A computational study. *Proc. Natl. Acad. Sci. USA*, *102*, 7002–7007.

Boudreau, C. E., Williford, T. H., & Maunsell, J. H. R. (2006). Effects of task difficulty and target likelihood in area V4 of macaque monkeys. *J. Neurophysiol. 96*, 2377–2387.

Butts, D. A., & Goldman, M. S. (2006). Tuning curves, neuronal variability, and sensory coding. *PLoS Biol., 4*, e92.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, *291*, 1560–1563.

Fries, P., Womelsdorf, T., Oostenveld, R., & Desimone, R. (2008). The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4. *J. Neurosci., 28*, 4823–4835.

Gruber, T., Müller, M. M., Keil, A., & Elbert, T. (1999). Selective visual-spatial attention alters induced gamma band responses in the human EEG. *Clin. Neurophysiol.*, *110*, 2074–2085.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Natl. Acad. Sci. USA*, *95*, 781–787.

Ikeda, T., & Hikosaka, H. (2003). Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron*, *39*, 693–700.

Jensen, O., Kaiser, J., & Lachaux, J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends Neurosci.*, *30*, 317–324.

Kanwisher, N. G., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nat. Rev. Neurosci.*, *1*, 91–100.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *J. Exp. Psychol. Hum. Percept. Perform.*, *21*, 451–468.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Percept. Psychophys.*, *56*, 183–197.

Lennie, P. (2003).The cost of cortical computation. *Curr. Biol.*, *13*, 493–497.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.*, *77*, 24–42.

Luck, S. J., Hillyard, S. A., Mangun, G. R., & Gazzaniga, M. S. (1989). Independent hemispheric attentional systems mediate visual search in split-brain patients. *Nature*, *342*, 543–545.

Mangun, G. R., & Hillyard, S. A. (1987). The spatial allocation of visual attention as indexed by event-related brain potentials. *Hum. Factors*, *29*, 195–211.

Mangun, G. R., & Hillyard, S. A. (1990). Allocation of visual attention to spatial locations: Tradeoff functions for event-related brain potentials and detection performance. *Percept. Psychophys.*, *47*, 532–550.

Martinez-Trujillo, J. C., & Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron*, *35*, 365–370.

Maunsell, J. H. R. (2004). Neuronal representations of cognitive state: Reward or attention? *Trends Cogn. Sci.*, *8*, 261–265.

Maunsell, J. H. R., & Cook, E. P. (2002). The role of attention in visual processing. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, *357*, 1063–1072.

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci*, *29*, 317–322.

McAdams, C. J., & Maunsell, J. H. R. (1999a). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.*, *19*, 431–441.

McAdams, C. J., & Maunsell, J. H. R. (1999b). Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron*, *23*, 765–773.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron*, *55*, 131–141.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*, 782–784.

Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.*, *14*, 2178–2189.

Müller, M. M., & Gruber, T. (2001). Induced gamma-band responses in the human EEG are related to attentional information processing. *Visual Cogn.*, *8*, 579–592.

Müller, M. M., Gruber, T., & Keil, A. (2000). Modulation of induced gamma band activity in the human EEG by attention and visual information processing. *Int. J. Psychophysiol.*, *38*, 283–299.

Navalpakkam, V., & Itti, L. (2007). Search goal tunes visual features optimally. *Neuron*, *53*, 605–617.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2003). Neuroimaging studies of attention: from modulation of sensory processing to top-down control. *J. Neurosci.*, *23*, 3990–3998.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*, 233–238.

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.*, *32*, 3–25.

Pouget, A., Deneve, S., & Ducom, J. C. (1999). Narrow versus wide tuning curves: What's best for a population code? *Neural Comput.*, *11*, 85–90.

Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, *26*, 703–714.

Roesch, M. R., & Olson, C. R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J. Neurophysiol.*, *90*, 1766–1789.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci.*, *18*, 3870–3896.

Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *J. Neurosci.*, *21*, 1340–1350.

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is consistent with temporal integration of random EPSPs. *J. Neurosci.*, *13*, 334–350.

Sparks, D. L. (1999). Conceptual issues related to the role of the superior colliculus in the control of gaze. *Curr. Opin. Neurobiol.*, *9*, 698–707.

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, *240*, 338–340.

Spitzer, H., & Richmond, B. J. (1991). Task difficulty: Ignoring, attending to, and discriminating a visual stimulus yield progressively more activity in inferior temporal neurons. *Exp. Brain Res.*, *83*, 340–348.

Steinmetz, P. N., Roy, A., Fitzgerald, P. J., Hsiao, S. S., Johnson, K. O., & Niebur, E. (2000). Attention modulates synchronized firing in primate somatosensory cortex. *Nature*, *404*, 187–190.

Taylor, K., Mandon, S., Freiwald, W.A., & Kreiter, A. K. (2005). Coherent oscillatory activity in monkey area V4 predicts successful allocation of attention. *Cereb. Cortex*, *15*, 1424–1437.

Tiesinga, P., Fellous, J. M., & Sejnowski, T. J. (2008). Regulation of spike timing in visual cortical circuits. *Nat. Rev. Neurosci.*, *9*, 97–107.

Tiesinga, P. H. E., Fellous, J. M., Salinas, E., Jose, J. V., & Sejnowski, T. J. (2004). Inhibitory synchrony as a mechanism for attentional gain modulation. *J. Physiol. Paris*, *98*, 296–314.

Treue, S., & Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*, 575–579.

Treue, S., & Maunsell, J. H. R. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J. Neurosci.*, *19*, 7591–7602.

Troyer, T. W., Krukowski, A. E., Priebe, N. J., & Miller, K. D. (1998). Contrast-invariant orientation tuning in cat visual cortex: Thalamocortical input tuning and correlation-based intracortical connectivity. *J. Neurosci.*, *18*, 5908–5927.

Verghese, P., & Pelli, D. G. (1992). The information capacity of visual attention. *Vis. Res.*, *32*, 983–995.

Williford, T., & Maunsell, J. H. R. (2006). Effects of spatial attention on contrast response functions in macaque area V4. *J. Neurophysiol.*, *96*, 40–54.

Womelsdorf, T., Fries, P., Mitra, P. P., & Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature*, *439*, 733–736.

Wyart, V., & Tallon-Baudry, C. (2008). Neural dissociation between visual awareness and spatial attention. *J. Neurosci.*, *28*, 2667–2679.

Yoshor, D., Ghose, G. M., Bosking, W. H., Sun, P., & Maunsell, J. H. R. (2007). Spatial attention does not strongly modulate neuronal responses in early visual cortex. *J. Neurosci.*, *27*, 13205–13209.

20 Selective Attention Through Selective Neuronal Synchronization

THILO WOMELSDORF AND PASCAL FRIES

ABSTRACT Selective attention relies on the dynamic restructuring of cortical information flow in order to prioritize neuronal communication from those neuronal groups conveying information about behaviorally relevant information, while reducing the influence from groups encoding irrelevant and distracting information. Electrophysiological evidence suggests that such selective neuronal communication is instantiated and sustained through selective neuronal synchronization of rhythmic gamma-band activity within and between neuronal groups: Attentionally modulated synchronization patterns evolve rapidly, are evident even before sensory inputs arrive, follow closely subjective readiness to process information in time, can be sustained for prolonged time periods, and carry specific information about top-down selected sensory features and motor aspects. These functional implications of selective synchronization patterns are complemented by recent insights about the mechanistic consequences of rhythmic synchronization, showing that selective neuronal interactions are subserved by neuronal synchronization that is selective in space, time, and frequency.

Top-down attention is the key mechanism to restructure cortical information flow in order to prioritize processing of behaviorally relevant over irrelevant and distracting information (Gilbert & Sigman, 2007). The behavioral consequences of attentional restructuring of information flow are manifold. Attended sensory inputs are processed more rapidly and accurately and with higher spatial resolution and sensitivity for fine changes, while nonattended information appears lower in contrast and is sometimes not perceived at all (Carrasco, Ling, & Read, 2004; Simons & Rensink, 2005).

These functional consequences of attention require temporally dynamic and selective changes of neuronal interactions spanning multiple levels of neuronal information processing: Attentional selection (1) modulates interactions among single neurons within cortical microcircuits, (2) modulates the impact of selective local neuronal groups conveying relevant information within functionally specialized brain areas, and (3) controls long-range interactions among neuronal groups from distant brain areas (Maunsell & Treue, 2006; Mitchell, Sundberg, & Reynolds, 2007; Reynolds & Chelazzi, 2004; Womelsdorf & Fries, 2007).

For all these levels of neuronal interactions, converging evidence suggests that the selective modulation of interactions critically relies on selective synchronization. Neuronal synchronization is typically oscillatory in nature; that is, neurons fire and pause together in a common rhythm. When synchronization is rhythmic, it is often addressed as coherence, and we will use these terms interchangeably. This rhythmic synchronization can influence neuronal interactions in several ways: (1) Spikes that are synchronized will have a larger impact on a target neuron than spikes that are not synchronized (Azouz & Gray, 2003; Salinas & Sejnowski, 2001). (2) Local inhibition that is rhythmically synchronized leaves periods without inhibition, while nonsynchronized inhibition will prevent local network activity continuously (Tiesinga, Fellous, Salinas, Jose, & Sejnowski, 2004). (3) Rhythmic synchronization of a local group of neurons will modulate the impact of input to that group, and therefore the impact of rhythmic input will depend on the synchronization between input and target (Womelsdorf et al., 2007). These mechanisms are at work on all levels of attentional selection: At the level of microcircuits, inhibitory interneuron networks have been shown to impose rhythmic synchronization capable of effectively controlling the gain of the neuronal spiking output (Bartos, Vida, & Jonas, 2007; Tiesinga, Fellous, & Sejnowski, 2008). At the level of local neuronal groups, attention is known to selectively synchronize the responses of those neurons conveying information about the attended feature or location (Womelsdorf & Fries, 2007). And the coherent output from these local neuronal groups has been shown to selectively synchronize over long-range connections with task-relevant neuronal groups in distant brain regions (Buschman & Miller, 2007; Saalmann,

THILO WOMELSDORF and PASCAL FRIES Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

Pigarev, & Vidyasagar, 2007; Schoffelen, Oostenveld, & Fries, 2005; Sejnowski & Paulsen, 2006).

These empirical insights suggest that mechanisms underlying neuronal synchronization could be the primary target of selective attention. In particular, top-down attention may act by biasing rhythmic synchronization to establish and sustain a selective neuronal communication structure (Fries, 2005). In the following, we begin by outlining this conceptual framework for selective attention through selective synchronization. We then survey basic insights from empirical and theoretical studies suggesting that rhythmic synchronization is particularly suited to control the selective routing of neuronal information flow, and we review how attention recruits these mechanisms across all levels of cortical processing.

Attentional selection as a dynamic instantiation of a selective neuronal communication structure

During natural sensation, top-down control is dynamically established during ongoing processing. Experimentally, top-down signals are set by task instructions, as well as by instructional cues defining relevant and irrelevant sensory features of the input stream during task performance. In typical paradigms of selective attention, the sensory input is kept identical across trials with variations only in covert attention to different aspects of that input. In such tasks, neuronal responses are modulated with rapid temporal dynamics and high spatial selectivity throughout the cerebral cortex (figure 20.1*A*).

The temporal dynamics of attentional selection are illustrated by recent evidence of a rapid onset of selective neuronal response modulation in cortical areas as far apart as frontal cortex and primary visual cortices in the macaque brain (Khayat, Spekreijse, & Roelfsema, 2006; Monosov, Trageser, & Thompson, 2008). In these studies, monkeys were instructed to detect a predefined target stimulus in visual displays to guide saccadic eye movement. In frontal and parietal cortex, attentional selection occurred within the first 120 ms following the sensory onset of target and distracter stimuli, allowing prediction of the spatial focus of attention (Gottlieb, 2002; Monosov et al.). Already about 30 ms later, top-down information changes neuronal responses at the earliest visual cortical processing stage in primary visual cortex (Khayat et al.; Roelfsema, Tolboom, & Khayat, 2007), causing a response enhancement for stimuli overlapping with the attentional target stimulus. These findings demonstrate that top-down control restructures cortical activity to sensory inputs across distant cortical sites on a rapid time scale. Attention amplifies almost instantaneously (i.e., with the sensory response latency) the influence of local groups of neurons conveying behaviorally relevant information and attenuates the influence of neuronal groups coding for irrelevant inputs. This finding suggests that distributed groups that process "attended" inputs also interact effectively, establishing selective neuronal communication structures on top of the existing infrastructure of anatomical connections (figure 20.1*A*): Interactions among neurons conveying information about attended locations or features are rendered effective, while anatomical connections between neuronal groups activated by distracting information are rendered ineffective.

Beyond the temporal dynamics of attentional selection, its spatial selectivity in restructuring cortical information flow is particularly evident across successive processing stages in visual cortex. Neurons at the highest visual processing stage in inferior temporal (IT) cortex have receptive fields that span much of a visual field and respond selectively to complex objects composed of simpler visual features. Part of this selectivity arises from their broad and convergent anatomical input from neurons in earlier processing stages having smaller receptive fields and simpler tuning properties. During natural vision, the large receptive field of an IT neuron will typically contain multiple objects. However, when attention is directed to only one of those objects, the IT neuron response is biased toward the response that would be obtained if only the attended object were presented (Chelazzi, Miller, Duncan, & Desimone, 1993; Moran & Desimone, 1985; Sejnowski & Paulsen, 2006; Sheinberg & Logothetis, 2001).

Such dynamic biasing of responses in IT cortex could be achieved by selective enhancement (suppression) of the impact of those afferent inputs from neurons in earlier visual areas coding for the attended (nonattended) input (Reynolds, Chelazzi, & Desimone, 1999). However, the mechanisms underlying this up- and down-modulation of input gain for subsets of converging connections are only poorly understood, but likely entail a selective increase of temporally precise and coincident inputs from neurons that are activated by attended input in earlier areas. This relevance of spike timing is suggested by fine-grained attentional modulation of precise neuronal synchronization within area V4 (Bichot, Rossi, & Desimone, 2005; Taylor, Mandon, Freiwald, & Kreiter, 2005; Womelsdorf, Fries, Mitra, & Desimone, 2006). Enhanced synchronization of the spiking output among neuronal groups that are activated by attended sensory input (Fries, Womelsdorf, Oostenveld, & Desimone, 2008) results in enhanced coincident arrival of their spikes at their postsynaptic target neurons in area IT. Temporally coincident input is highly effective in driving neuronal activity (Azouz & Gray, 2003; Salinas & Sejnowski, 2001; Tiesinga et al., 2008). It is therefore likely that selective synchronization within area V4 underlies attentional biasing within IT cortex and could thus underlie effective spatial routing of information flow within visual cortex.

Please note that neuronal synchronization is in principle independent of firing rate, in terms of both metrics

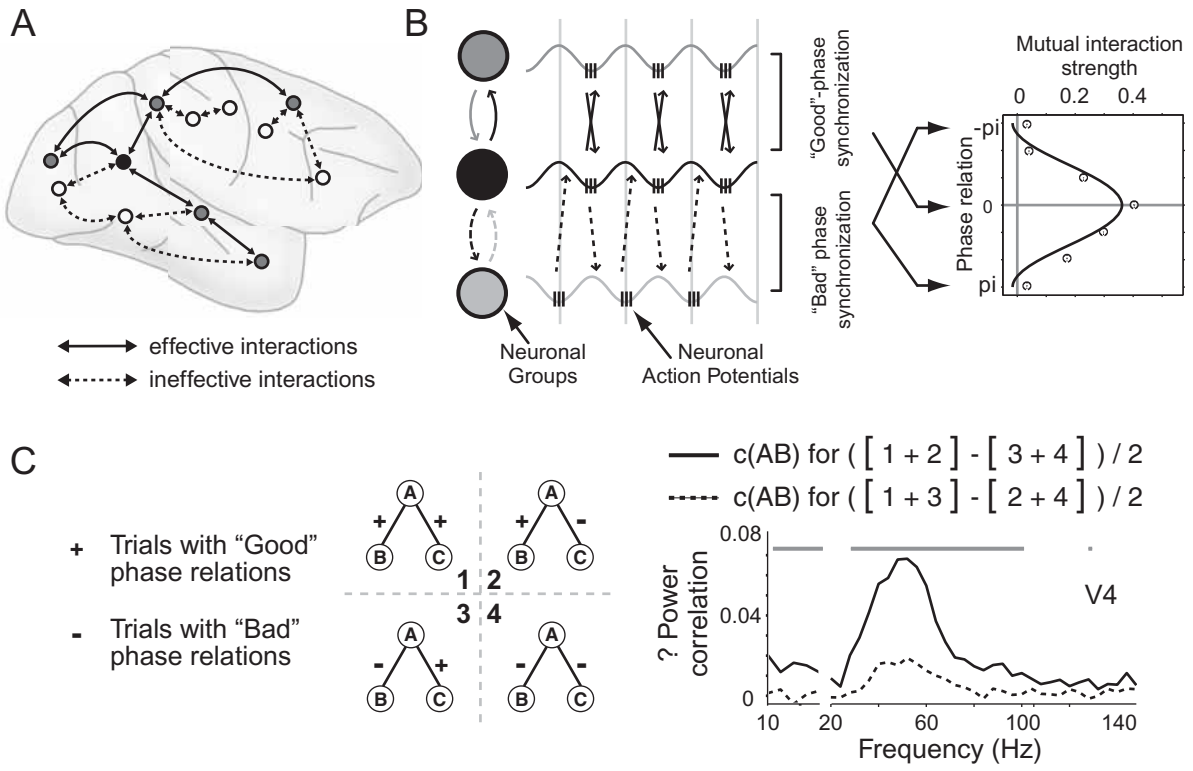


FIGURE 20.1 Selective synchronization renders neuronal interactions among subsets of neuronal groups effective. (*A*) Anatomical connectivity (sketched as lines) provides a rich infrastructure for neuronal communication among neuronal groups (circles) throughout the cortex. With selective attention, only a small subset of these connections are rendered effective (solid lines). Interactions among groups conveying irrelevant information (light gray circles) for the task at hand are rendered less effective (dashed lines). (*B*) Illustration of the hypothesized role of selective synchronization for selective communication among three neuronal groups (circles). Rhythmic activity (local field potential, or LFP, oscillations with spikes in troughs) provide briefly recurring time windows of maximum excitability (LFP troughs), which are either in phase (black and dark gray groups), or in antiphase (black and light gray groups). The plot on the right shows that mutual interactions (upper axis, correlation of the power of the LFP and the neuronal spiking response between neuronal groups) are high during periods of in-phase synchronization and lower otherwise. (*C*) The trial-by-trial interaction pattern between neuronal groups (A to B and A to C) is predicted by the pattern of synchronization: If AB synchronizes at a good phase, their interaction is strongest, irrespective of whether A synchronizes with C at good or bad phase relations in the same trials. Thus the spatial pattern of mutual interactions can be predicted by the phase of synchronization among rhythmically activated neuronal groups. (Panels in *B* and *C* adapted from Womelsdorf et al., 2007.)

and physiology. The different metrics used for quantifying synchronization are typically normalized for firing rate. Physiologically, there are examples where enhanced firing rates are associated with strongly reduced synchronization, for example, the stimulus-induced alpha-band desynchronization in the superficial layers of monkey V4 (Fries et al., 2008). Neuronal gamma-band synchronization typically emerges when neuronal groups are activated, and therefore it is in most cases associated with increased firing rates. However, firing rates and gamma-band synchronization can also be dissociated from each other, and this pattern can be found primarily when firing rate changes are driven not by changes in bottom-up input (e.g., stimulus changes) but rather by changes in top-down input (e.g., attention or stimulus selection) (Fries, Schröder, Roelfsema, Singer, & Engel, 2002; Womelsdorf et al., 2006).

Synchronization is a neuronal population phenomenon, and it is often very difficult to assess it with recordings from isolated single units. Correspondingly, many studies of neuronal synchronization use recordings of multiunit activity and/or of the local field potential (LFP). The LFP reflects the summed transmembrane currents of neurons within a few hundred micrometers of tissue. Since synchronized currents sum up much more efficiently than unsynchronized currents, the LFP primarily reflects synchronized synaptic activity. Changes in LFP power typically correlate very well with changes in direct measures of neuronal synchronization.

Rhythmic synchronization within a neuronal group not only increases its impact on postsynaptic target neurons in a feedforward manner. It also rhythmically modulates the group's ability to communicate, such that rhythmic synchronization between two neuronal groups likely subserves their interaction, because rhythmic inhibition within the two

groups is coordinated and mutual inputs are optimally timed. We capture these implications in the framework of selective attention through selective synchronization (Fries, 2005).

Selective attention through selective synchronization

Local neuronal groups frequently engage in periods of rhythmic synchronization. During activated states, rhythmic synchronization is typically evident in the gamma frequency band (30–90 Hz) (Engel, Konig, Gray, & Singer, 1990; Gray, Konig, Engel, & Singer, 1989; Hoogenboom, Schoffelen, Oostenveld, Parkes, & Fries, 2006). In vitro experiments and computational studies suggest that gamma-band synchronization emerges from the interplay of excitatory drive and rhythmic inhibition imposed by interneuron networks (Bartos et al., 2007; Börgers, Epstein, & Kopell, 2005; Börgers & Kopell, 2003; Buia & Tiesinga, 2006). Interneurons impose synchronized inhibition onto the local network (Bartos et al.; Hasenstaub et al., 2005; Vida, Bartos, & Jonas, 2006). The brief time periods between inhibition provide time windows for effective neuronal interactions with other neuronal groups, because they reflect enhanced postsynaptic sensitivity to input from other neuronal groups, as well as maximal excitability for generating spiking output to other neuronal groups (Azouz, 2005; Azouz & Gray, 2003; Fries, Nikolic, & Singer, 2007; Tiesinga et al., 2008). As a consequence, when two neuronal groups open their temporal windows for interaction at the same time, they will be more likely to mutually influence each other (Womelsdorf et al., 2007). The consequences for selective neuronal communication are illustrated in figure 20.1*B*: If the rhythmic synchronization within neuronal groups is precisely synchronized between the two groups, they are maximally likely to interact. By the same token, if rhythmic activity within neuronal groups is uncorrelated between groups or synchronizes consistently out of phase, their interaction is curtailed (figure 20.1*B*).

This scenario entails that the pattern of synchronization between neuronal groups flexibly structures the pattern of interactions between neuronal groups (figure 20.1*C*). Consistent with this hypothesis, the interaction pattern of one neuronal group (A) with two other groups (B and C) can be predicted by their pattern of precise synchronization (figure 20.1*C*). This result has recently been demonstrated for interactions of triplets of neuronal groups from within and between areas in awake cat and monkey visual cortex (Womelsdorf et al., 2007). This study measured the trial-by-trial changes in correlated amplitude fluctuation and changes in precise synchronization between pair AB and pair AC, using the spontaneous variation of neuronal activity during constant visual stimulation. The strength of amplitude covariation—that is, the covariation of power in the LFP and/or multiunit spiking responses—was considered the measure of mutual interaction strength. The results showed that the interaction strength of AB could be inferred from the phase of gamma-band synchronization between group A and group B, being rather unaffected by the phase of synchronization of group A with group C (figure 20.1*C*). This finding was evident for triplets of neuronal groups spatially separated by as little as 650 μm, illustrating a high spatial resolution and specificity of the influence of precise phase synchronization between neuronal groups on the efficacy of neuronal interaction. Importantly, additional analysis supported a mechanistic role for the phase of synchronization between rhythmic activities to modulate the effective interaction strength (Womelsdorf et al.). In particular, precise phase synchronization preceded higher amplitude covariations in time by a few milliseconds, arguing for a causal influence of precise phase synchronization to trigger neuronal interactions. Taken together, these results provide the most direct evidence available so far to suggest a critical mechanistic role of selective synchronization for neuronal interactions. They demonstrate that synchronization patterns can shape neuronal interactions with high specificity in time, space, and frequency.

Importantly, these same characteristics of selective neuronal interactions are the key elements underlying selective attention. Attentional selection dynamically evolves at a rapid time scale and with high spatial resolution by enhancing (reducing) the effective connectivity among neuronal groups conveying task relevant (irrelevant) information. Such dynamic restructuring of neuronal interactions could be accomplished through mechanisms evoking selective synchronization patterns within interneuron networks. Selective changes of precise synchronization in local neuronal groups are capable of modulating in a self-emergent manner selective interaction patterns across neuronal groups (Börgers & Kopell, 2008; Mishra, Fellous, & Sejnowski, 2006; Tiesinga et al., 2008).

The outlined scheme of selective attention implemented as selective neuronal synchronization comprises explicit assumptions that selective attention affects interneuron networks and synchronization patterns during task performance. In the following we survey the available insights on interneuron networks and review the emerging signatures of attentional modulation of selective synchronization patterns in macaque cortex.

Synchronization in interneuron networks and their attentional modulation

Interneurons comprise about a fifth of the neuron population, but despite their ubiquitous presence, their functional roles underlying cortical computations or cognitive processes are far from understood (Markram et al., 2004). However, a central role for the control of local cortical network activity has been suggested for a large class of interneurons (primarily of the basket cell type) (Buzsaki, 2006). These neurons

target perisomatic regions of principal cells and are thereby capable of determining the impact of synaptic inputs arriving at sites distal to a cell's soma. Such perisomatic connectivity could therefore critically control the input gain of principal cells across a large population of principal cells (Buzsaki, Kaila, & Raichle, 2007; Cobb, Buhl, Halasy, Paulsen, & Somogyi, 1995; Markram, Wang, & Tsodyks, 1998; Rudolph, Pospischil, Timofeev, & Destexhe, 2007; Tiesinga, Fellous, et al., 2004). As described in the previous paragraph, the inhibitory synaptic influence is inherently rhythmic at high frequencies, carrying stronger gamma-band power than pyramidal cells (Bartos et al., 2007; Hasenstaub et al., 2005).

The prominent role of these high-frequency inputs in shaping the spiking output of principal cells has recently been demonstrated directly in visual cortex of the awake cat. It was shown that the spiking of principal cells is indeed preceded by brief periods of reduced inhibition (Rudolph et al., 2007; see also figure 8 of Hasenstaub et al., 2005). Taken together, these findings suggest that interneurons are the source of rhythmic inhibition onto a local group of neurons synchronizing the discharge of pyramidal cells to the time windows between inhibition.

In the context of selective attention, interneuron networks could be activated by various possible sources. They may be activated by transient and spatially specific neuromodulatory inputs (Lin, Gervasoni, & Nicolelis, 2006; Rodriguez, Kallenbach, Singer, & Munk, 2004). Alternatively, selective attention could target local interneuron networks directly by way of top-down inputs from neurons in upstream areas (Buia & Tiesinga, 2008; Mishra et al., 2006; Tiesinga et al., 2008). In these models, selective synchronization emerges either by depolarizing selective subsets of interneurons (Buia & Tiesinga, 2008; Tiesinga & Sejnowski, 2004) or by biasing the phase of rhythmic activity in a more global inhibitory interneuron pool (Mishra et al., 2006). In either case, rhythmic inhibition controls the spiking responses of groups of excitatory neurons, enhancing the impact of neurons spiking synchronously within the periods of disinhibition, while actively reducing the impact of neurons spiking asynchronously to this rhythm. This suppressive influence on excitatory neurons, which are activated by distracting feedforward input, reflects the critical ingredient for the concept of selective attention through selective synchronization: Attention not only enhances synchronization of already more coherent activity representing attended stimuli, but also actively suppresses the synchronization and impact of groups of neurons receiving strong, albeit distracting, inputs, because they arrive at nonoptimal phase relations to the noninhibited periods in the target group. The computational feasibility of both facilitatory and suppressive aspects and the critical role of the timing of inhibitory circuits have recently received direct support (Börgers & Kopell, 2008).

Despite the prominent computational role of interneuron activity for selective communication, there are only sparse insights into their implications in selective information processing during cognitive task performance. The basic prediction from the preceding models is that interneurons are strongly attentionally modulated. Consistent with this presupposition, a recent study by Mitchell, Sundberg, and Reynolds (2007) reports a clear attentional modulation of putative interneurons in visual area V4 during a selective attention task requiring monkeys to track moving grating stimuli (Mitchell et al.). Putative interneurons showed similar relative increases in firing rate and greater increases in reliability compared to putative pyramidal neurons. However, tests of more refined predictions about the relative modulation of synchronization and the phase relation of spiking responses of inhibitory and excitatory neuron types still need to be conducted (Buia & Tiesinga, 2008).

Selective modulation of synchronization during attentional processing

Direct evidence for the functional significance of selective synchronization within local neuronal groups for attentional selection has been obtained from recordings in macaque visual cortical area V4 (Womelsdorf & Fries, 2007). One consistent result across studies is that spatial attention enhances gamma-band synchronization within neuronal groups that have receptive fields overlapping the attended location (Fries, Reynolds, Rorie, & Desimone, 2001; Taylor et al., 2005; Womelsdorf et al., 2006). The enhanced rhythmic synchronization is strongly evident within the LFP signal, which is a compound signal of activity within a local neuronal group, and is likewise reflected in more precise synchronization of neuronal spiking responses to the LFP. Importantly, the synchronization among the spiking output from neurons coding for the attended location is also enhanced compared to the spiking output of neurons activated by a nonattended distracter stimulus (figure 20.2) (Fries et al., 2008). These attentional effects on spike-to-spike synchronization imply that the postsynaptic targets receive more coherent input from the neuronal groups that convey behaviorally relevant information.

Functional Implications of Selective Gamma-Band Synchronization In addition to the described attentional effect, recent studies demonstrated that the precision of local synchronization in visual area V4 is closely related to task performance, including behavioral accuracy and the time to detect behaviorally relevant stimulus changes (Taylor et al., 2005; Womelsdorf et al., 2006). This conclusion was derived from an error analysis of the pattern of synchronization in area V4 (Taylor et al.). In this study, the spatial focus of attention could be inferred from the pattern of

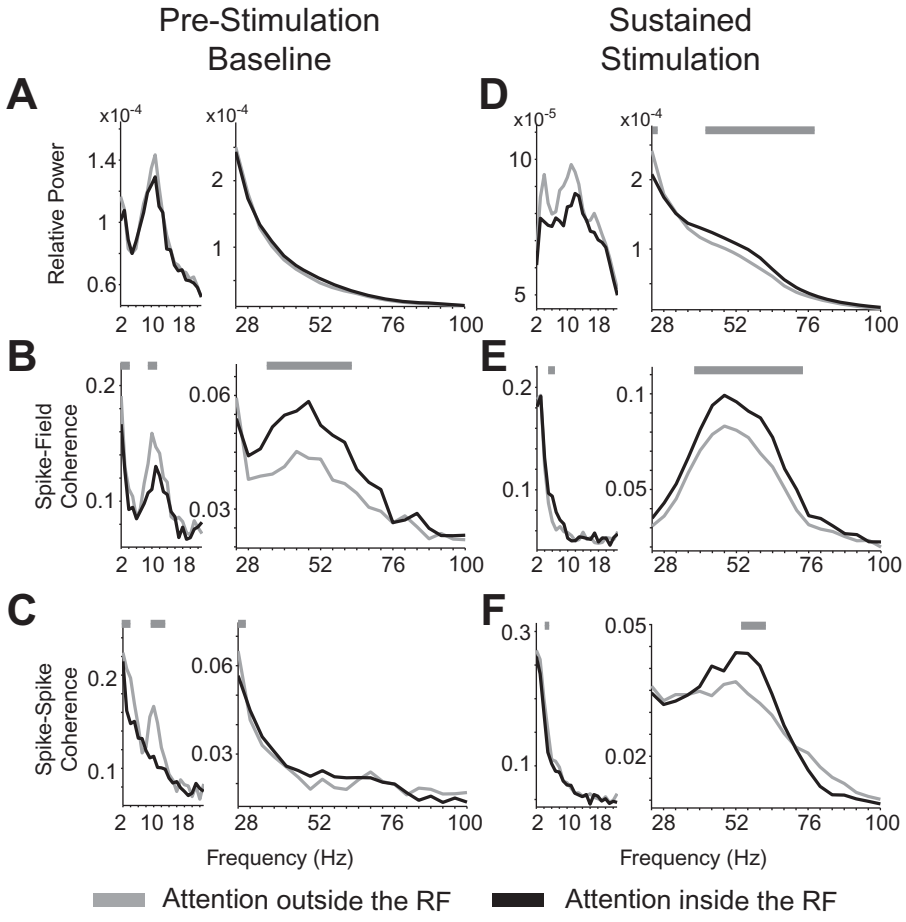


FIGURE 20.2 The pattern of attentional modulation of synchronization in macaque visual area V4 before and during sensory stimulation. (*A–C*) Attentional modulation of relative LFP power (*A*), spike-to-LFP coherence (*B*), and spike-to-spike coherence (*C*) across low and high frequencies during the baseline period of a spatial attention task. Monkeys either attended (dark lines) or ignored (gray lines) the receptive field location of the recorded neuronal groups in blocks of trials. (*D–F*) Attentional modulation of the neuronal response during stimulation with an attended/ignored moving grating. Same format as in *A–C*. Horizontal gray bars denote frequencies with significant attentional effects. (Adapted from Fries, Womelsdorf, Oostenveld, & Desimone, 2008.)

synchronization measured through epidural electrodes. Gamma-band synchronization was not only stronger for correct trials than for miss trials, but additionally, the degree of synchronization predicted whether the monkey was paying attention to the distracter. Thus this study demonstrated that gamma-band synchronization reflects the actual allocation of attention rather than merely the attentional cuing itself. Furthermore, another recent study demonstrated that the precision of stimulus-induced gamma-band synchronization predicts how rapidly a stimulus change can be reported behaviorally. When monkeys were spatially cued to select one of two stimuli in order to detect a color change of the attended stimulus, the speed of change detection could be partly predicted by the strength of gamma-band synchronization shortly before the stimulus change actually occurred (Womelsdorf et al., 2006). Importantly, the reaction times to the stimulus change could not be predicted at times before the stimulus change by overall firing rates, nor by synchronization outside the gamma band. Notably, the correlation of gamma-band synchronization with the speed of change detection showed high spatial selectivity: Neurons activated by an unattended stimulus engaged in lower synchronization when the monkeys were particularly fast in responding to the stimulus change at locations outside their receptive field. This finding rules out a possible influence of globally increased synchronization during states of enhanced alertness and arousal (Herculano-Houzel, Munk, Neuenschwander, & Singer, 1999; Munk, Roelfsema, Konig, Engel, & Singer, 1996; Rodriguez et al., 2004). And it argues for a fine-grained influence of synchronization to modulate the effective transmission of information about the stimulus change to postsynaptic target areas concerned with the planning and execution of responses.

These behavioral correlates of gamma-band synchronization during selective attention tasks are complemented by a variety of correlational results linking enhanced gamma-band synchronization to efficient task performance in various

paradigms involving attentional processing. For example, in memory-related structures the strength of gamma-band synchronization has been linked to the successful encoding and retrieval of information (Montgomery & Buzsaki, 2007; Sederberg et al., 2006; Sederberg, Kahana, Howard, Donner, & Madsen, 2003; Sederberg et al., 2007).

Selective Gamma-Band Coherence Beyond Visual Cortex These results of selective gamma-band synchronization with selective spatial attention are supported by a growing number of converging findings from human EEG and MEG studies (Doesburg, Roggeveen, Kitajo, & Ward, 2008; Fan et al., 2007; Gruber, Müller, Keil, & Elbert, 1999; Landau, Esterman, Robertson, Bentin, & Prinzmetal, 2007; Wyart & Tallon-Baudry, 2008). Importantly, attention modulates gamma-band synchronization beyond sensory visual cortex. It has been reported for auditory cortex (Kaiser, Hertrich, Ackermann, & Lutzenberger, 2006; Tiitinen et al., 1993) and more recently in somatosensory cortex. Spatial attention for tactile discrimination at either the right or left index finger in humans enhanced stimulus-induced gamma-band synchronization in primary somatosensory cortex when measured with MEG (Bauer, Oostenveld, Peeters, & Fries, 2006; Hauck, Lorenz, & Engel, 2007). Similar topographies and dynamics of gamma-band synchronization were shown to correlate with the actual perception of somatosensory induced pain (Gross, Schnitzler, Timmermann, & Ploner, 2007). Importantly, enhanced oscillatory dynamics in the gamma band during tactile perception is not restricted to the somatosensory cortex (Ohara, Crone, Weiss, & Lenz, 2006). In recent intracranial recordings in humans, synchronization was modulated across somatosensory cortex, medial prefrontal, and insular regions when subjects had to direct attention to painful tactile stimulation (Ohara et al.).

Spatially Specific Synchronization Patterns During Preparatory Attentional States The described gamma-band modulation of rhythmic activity is most prominent during activated states. However, attentional top-down control biases neuronal responses in sensory cortices already before sensory inputs impinge on the neuronal network (Fries, Reynolds, et al., 2001; Fries et al., 2008; Luck, Chelazzi, Hillyard, & Desimone, 1997). In many attention studies, the instructional cue period is followed by a temporal delay void of sensory stimulation. During these preparatory periods, top-down signals set the stage for efficient processing of expected stimulus information, rendering local neuronal groups ready to enhance the representation of attended sensory inputs. Intriguingly, the described preparatory bias is evident in selective synchronization patterns in the gamma band and in rhythmic synchronization at lower frequencies.

In macaque visual cortical area V4, neurons synchronized their spiking responses to the LFP in the gamma band more precisely when monkeys expected a target stimulus at the receptive field location of the respective neuronal group (figure 20.2*B*). This modulation was evident even though rhythmic activity proceeded at far lower levels in the absence of sensory stimulation compared to synchronization strength during high-contrast sensory drive. Lower overall strength, and correspondingly lower signal-to-noise ratio, may account for the lack of significant gamma-band modulation of LFP power or spike-to-spike synchronization during the prestimulus period when compared to attentional modulation during stimulation (figure 20.2).

During preparatory periods, and thus in the absence of strong excitatory drive to the local network, rhythmic activity is dominated by frequencies lower than the gamma band. In the described study from macaque V4, prestimulus periods were characterized by alpha-band peaks of local rhythmic synchronization when monkeys attended away from the receptive location of the neuronal group. Figure 20.2*B,C* demonstrates reduced locking of neuronal spiking in the alpha band to the LFP and to spiking output of nearby neurons (figure 20.2*B,C*). This finding is in general agreement with various studies demonstrating reduced alpha-band activity during attentional processing (Bauer et al., 2006; Pesaran, Pezaris, Sahani, Mitra, & Andersen, 2002; Rihs, Michel, & Thut, 2007; Worden, Foxe, Wang, & Simpson, 2000; Wyart & Tallon-Baudry, 2008). Interestingly, human EEG studies extend this finding by showing that the degree of alpha-frequency desynchronization during prestimulus intervals of visuospatial attention tasks indicates how fast a forthcoming target stimulus is processed (Jin, O'Halloran, Plon, Sandman, & Potkin, 2006; Sauseng et al., 2006; Thut, Nietzel, Brandt, & Pascual-Leone, 2006). For example, reaction times to a peripherally cued target stimulus are partially predicted by the lateralization of alpha activity in the one-second period before target appearance (Thut et al.). While this predictive effect was based predominantly on reduced alpha-band responses over the hemisphere processing the attended position, recent studies suggest that alpha-band oscillations are selectively enhanced within local neuronal groups processing distracting information, that is, at unattended locations (Kelly, Lalor, Reilly, & Foxe, 2006; Rihs et al.; Yamagishi et al., 2003). These findings suggest that rhythmic alpha-band synchronization may play an active role in preventing the signaling of stimulus information. According to this hypothesis, attention is thought to up-regulate alpha-band activity of neuronal groups expected to process distracting stimulus information, rather than to down-regulate local alpha-band synchronization for neuronal groups processing attended stimulus features and locations.

Synchronization Patterns Reflecting Temporal Expectancies of Target Processing The previous paragraph surveyed evidence for an influence of spatially specific expectancy of target and distracter stimuli on synchronization patterns in visual cortex. In addition to spatially selective expectancy, the expectation of the occurrence of behaviorally relevant target events is known to influence neuronal synchronization patterns and firing rates in parietal and frontal cortex (Ghose & Maunsell, 2002; Janssen & Shadlen, 2005; Riehle, 2005; Schoffelen et al., 2005). Attentional modulation of neuronal firing rates in extrastriate visual area MT is strongest around the time point at which the subjective anticipation for a target change, given that it had not occurred before in the trial (i.e., the hazard rate), is maximal (Ghose & Maunsell, 2002). In premotor and motor cortex, the hazard rate is smoothly reflected in the strength of synchronization (Riehle, 2005; Schoffelen et al.). Importantly, enhanced readiness to respond to attended sensory changes is thereby functionally closely linked to long-range synchronization of motor cortex with spinal motor units suggesting a direct mechanistic influence of synchronization on the speed to respond to behaviorally relevant sensory events (Schoffelen et al.).

An influence of temporal expectancy on synchronization in early sensory cortices has recently been demonstrated in recordings in primary visual cortex of macaques (Lakatos, Karmos, Metha, Ulbert, & Schroeder, 2008). In this study, monkeys were cued to detect deviant sensory stimuli in either an auditory or visual input stream to receive a reward. Auditory and visual stimuli alternated, and both stimulus streams followed a noisy 1.55-Hz rhythm. This low-frequency rhythm of sensory inputs entrained neuronal responses in early visual cortex, such that responses to individual stimuli in the visual stream added to the entrained response. Attention to the visual stream amplified the entrainment (figure 20.3*A*), but the most prominent attentional effect was evident in the phase of the 1.55-Hz entrainment in the superficial layers of visual cortex: This entrainment was always determined by the stimulus stream that was attended; that is, it switched by half a cycle when attention switched from the visual to the auditory stream (figure 20.3*B*), which had a phase opposite to the visual stream. Importantly, low-frequency fluctuations in the LFP likely reflect fluctuations in neuronal excitability. With attention to the visual (auditory) input stream, the phase corresponding to maximal (minimal) neuronal excitability occurred around the average time when the target information was most likely to reach visual cortex. Consistent with a functional role of the entrained delta phase, the authors reported the strongest attentional enhancement of gamma-band synchronization in the LFP and spiking activity around this time (figure 20.3*C,D*), and showed that the detection of deviant visual stimuli was fastest (slowest) when the delta phase at stimulus onset corresponded to maximal (minimal) neuronal excitability (figure 20.3*E*).

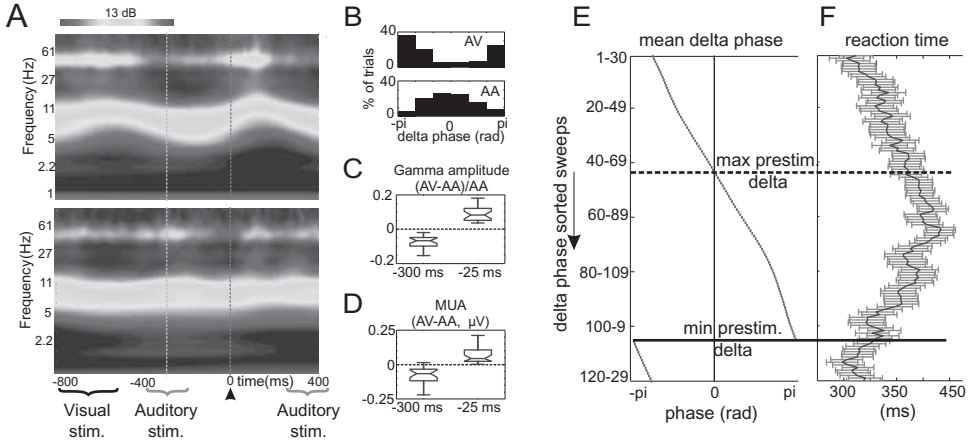


Figure 20.3 (*A*) Entrainment of synchronization from delta- to gamma-band frequencies in supragranular layers of the primary visual cortex during an auditory-visual change detection task. (*B*) Time-frequency spectrograms during attention to the visual (upper panel) and auditory (bottom panel) input stream, aligned to the onset time of the visual stimulus. The task cued monkeys to detect infrequent deviant stimuli in either the auditory (white noise tones) or visual (red light flashes) input stream. Visual (auditory) stimuli were onset at a regular interval of 650 ms ± 150 ms indicated below the time axis (mean stimulus rate, 1.55 Hz). (*C*) Entrainment of delta-frequency phase in visual cortex measured as the delta (1.55 Hz) phase at the time of visual stimulus onset when attention was directed to the visual (upper panel) and auditory (lower panel) modality across recording sessions. (*D, E*) Modulation of gamma band amplitude of the LFP (*D*) and multiunit activity (*E*) before and at visual stimulus onset. Positive values indicate enhancement with visual versus auditory attention. (*F*) Reaction times (*x*-axis) to the visual target stimulus sorted into groups of trials according to the prestimulus delta phase (at 0 ms to stimulus onset) (*y*-axis). Solid/dashed horizontal lines indicate the group of trials corresponding to maximum/minimum delta amplitudes. (Adapted from Lakatos, Karmos, Metha, Ulbert, & Schroeder, 2008.)

The described results suggest that top-down information selectively modulates excitability in early sensory cortices through changes in the phase of rhythmic entrainment in these areas. The exact frequency band underlying excitability modulations may extend from the low delta band directly imposed by the stimulus structure in the described study to the theta band around 4–8 Hz. This suggestion may be derived from the time-frequency evolution of LFP power in the theta band and its attentional modulation, shown in figure 20.3*A*. Intriguingly, similar to the effect of delta phase on the gamma-band response demonstrated directly in the discussed study (figure 20.3*B*), previous studies have linked the phase of rhythmic activity in the theta band to the strength of high-frequency gamma-band synchronization in rodent hippocampus and over large regions in the human cortex (Canolty et al., 2006; Csicsvari, Jamieson, Wise, & Buzsaki, 2003). An additional hint suggesting a functional relevance of low-frequency phase fluctuations can be found in a recent study in rodents demonstrating that neuronal spiking responses in rodent prefrontal cortex phase-lock to theta-band activity in the hippocampus during task epochs requiring spatial decisions in a working memory context (Jones & Wilson, 2005). In macaque visual cortex, the phase of theta-band synchronization has been directly linked to selective maintenance of task-related information (Lee, Simpson, Logothetis, & Rainer, 2005). Taken together, the emerging evidence demonstrates (1) that top-down, task-related information modulates low-frequency rhythmic activity, (2) that the phase of this rhythmic activity can be functionally related to task performance, and (3) that the phase of low-frequency activity shapes the strength of gamma-band synchronization in response to sensory inputs. As such, the pattern of selective synchronization in the gamma band described in the previous paragraphs could be tightly linked to underlying, selective low-frequency activity modulations. Whether both are coupled in an obligatory way, or whether the comodulation may be triggered by specific task demands, will be an interesting subject for future research.

Feature-Selective Modulation of Rhythmic Synchronization The preceding sections discussed evidence for selective neuronal synchronization patterns evolving with space-based attentional selection of sensory inputs. However, in addition to spatial selection, attention frequently proceeds only on top-down information about the behaviorally relevant sensory feature and independent of the exact spatial location at which input impinges on sensory cortices. Such feature-based attention is known to modulate the responses of neurons that are tuned to the attended feature such as a particular motion direction or the color of a visual stimulus (Maunsell & Treue, 2006).

Importantly, a recent study demonstrated that attention to a particular feature selectively synchronizes the responses of neurons tuned to the attended stimulus feature (Bichot, Rossi, & Disimone, 2005). In this study, spiking responses and LFPs were recorded from neuronal groups in macaque visual area V4 while monkeys searched in multistimulus displays for a target stimulus defined either by color, shape, or both. When monkeys searched, for example, for a red stimulus by shifting their gaze across stimuli on the display, the nonfoveal receptive fields of the recorded neurons could either encompass nontarget stimuli (e.g., of blue color) or the (red) target stimulus prior to the time when the monkey detected the target. The authors found that neurons synchronized to the LFP more strongly in response to their preferred stimulus feature when it was the attended search target feature rather than a distracter feature.

Thus attention enhanced synchronization of the responses of the neurons that shared a preference for the attended target feature—and irrespective of the spatial location of attention (Bichot et al., 2005). This feature-based modulation was also evident during a conjunction search task involving targets that were defined by two features: When monkeys searched for a target stimulus with a particular orientation and color (e.g., a red horizontal bar), neurons with preference to one of these features enhanced their neuronal synchronization (Bichot et al.). This enhancement was observed not only in response to the color-shape-defined conjunction target, but also in response to distracters sharing one feature with the target (e.g., red color). This latter finding corresponds well with the behavioral consequences of increased difficulty and search time needed for conjunction-defined targets.

This study shows that feature salience is indexed not only by changes in firing rates as has been shown before (Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999; Wannig, Rodriguez, & Freiwald, 2007), but also by selectively synchronizing neuronal responses depending on the similarity between neuronal feature preferences and the attended stimulus feature. The mechanisms behind this selective influence of featural top-down information could be based on a similar spatial weighting of interneuron network activity as implicated for spatial selection. Neuronal tuning to many basic sensory features is organized in regularly arranged local maps. Correspondingly, the tuning of groups of neurons measured with the LFP is locally highly selective. Importantly, neuronal stimulus preference is systematically related to the strength of neuronal synchronization in the gamma frequency band. This relationship has been demonstrated for stimulus orientation and spatial frequency (Frien, Eckhorn, Bauer, Woelbern, & Gabriel, 2000; Gray, Engel, Konig, & Singer, 1990; Kayser & König, 2004; Kreiter & Singer, 1996; Siegel & König, 2003), the speed and direction of visual motion (Liu & Newsome, 2006), and the spatial motor intentions and movement directions (Scherberger & Andersen, 2007; Scherberger, Jarvis, &

Andersen, 2005). These findings show that rhythmic synchronization conveys feature-selective information. Feature-based attention appears to recruit this property with high spatial resolution by modulating which neurons synchronize to the local rhythmic activity.

Taken together, the previous subsections surveyed the accumulating evidence demonstrating selective neuronal synchronization patterns that evolve with selective spatial and feature-based attention within sensory cortices. Only a few studies have extended these insights to ultimately reveal how synchronization patterns within sensory areas are related to selective neuronal interaction patterns between different visual areas and between visual and higher-order cortical areas during task performance. Recent evidence shows that such dynamic inter-areal interaction patterns are evident in long-range synchronization patterns between cortical areas.

Selective inter-areal synchronization during attentional processing

In the preceding sections, selective synchronization patterns evolved for local neuronal groups in sensory cortices and were evident primarily within a confined gamma frequency band. These findings critically support a functional role for gamma-band synchronization for the selective restructuring of neuronal communication during attentional processing (cf. figure 20.1). However, attentional processing relies on effective interactions between local subsets of neuronal groups from distant cortical regions. So far, only a few studies have investigated these inter-areal interaction patterns during task epochs with selective attention (Engel, Fries, & Singer, 2001; Varela, Lachaux, Rodriguez, & Martinerie, 2001; Womelsdorf & Fries, 2007). The emerging evidence from these studies points toward a critical role of rhythmic long-range synchronization in frequencies both within and below the gamma band, including prominently a beta band that spans frequencies from 15 Hz to 30 Hz.

Early studies in awake cats demonstrated transiently enhanced beta-frequency synchronization among visual cortical and premotor regions, and between visual cortex and thalamus during nonselective states of expectancy of a behaviorally relevant stimulus (in, e.g., go-no-go tasks) (Roelfsema, Engel, König, & Singer, 1997; von Stein, Chiang, & König, 2000; Wrobel, Ghazaryan, Bekisz, Bogdan, & Kaminski, 2007). Recent studies in the macaque monkey have extended these findings by showing that frontoparietal and intraparietal interactions between areas are accompanied by synchronization at high beta frequencies (20–35 Hz) during task epochs requiring searching for and selecting behaviorally relevant visual stimuli (Buschman & Miller, 2007; Saalmann et al., 2007). Figure 20.4 illustrates findings from a visual search task requiring monkeys to detect a search target that is either salient and pops out among distracting stimuli

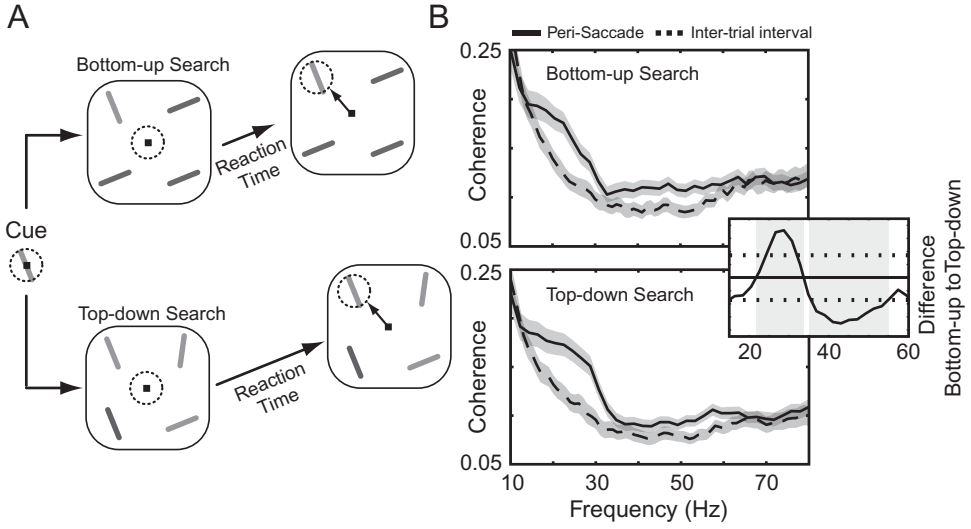


Figure 20.4 Selective modulation of long-range synchronization between frontal and parietal cortex during visual search. (*A*) Sketch of two visual search tasks used by Buschman and Miller (2007). A cue instructed monkeys about the orientation and color of a bar that was the later search target in a multistimulus display during a bottom-up search task (both target color and orientation were unique, upper panels) and during a top-down search task (target shared color or orientation with distracting stimuli, bottom panels). Monkeys covertly attended the multistimulus array and made a saccade to the target stimulus position as soon as they found it. (*B*) The authors measured the coherence of the LFP activity of neuronal groups in the frontal eye field and dorsolateral prefrontal cortex) and parietal area LIP. The line plots on the right show the coherence (*y*-axis) for different frequency bands (*x*-axis) in the bottom-up and top-down tasks, along with the coherence difference across tasks (solid line in inset). The results show that attentional demand modulated long-range frontoparietal coherence at different frequency bands. (Adapted from Buschman & Miller, 2007.)

("bottom-up search") or that is nonsalient because it shares features with distracting stimuli (Buschman & Miller, 2007). In contrast to bottom-up salient targets, the nonsalient target stimuli were detected more slowly, indicating that they require attentive search through the stimuli in the display before they are successfully detected ("top-down search"). Paralleling the difference in behavioral demands, the authors found a selective synchronization pattern among the LFPs in frontal and parietal cortex. While attentive "top-down search" enhanced specifically rhythmic synchronization at 20–35 Hz compared to the "bottom-up" search, the stimulus driven "bottom-up" search resulted in stronger inter-areal synchronization in the gamma-frequency band (figure 20.4*B*). The pattern of results is most likely due to relative differences in task demands in both search modes and was unaffected by differences in reaction times. Therefore these findings suggest that inter-areal communication during attentional top-down control is conveyed particularly through rhythmic synchronization in a high beta band, either in addition to or separate from the frequency of rhythmic interactions underlying bottom-up feedforward signaling.

Consistent with a functional role for top-down-mediated long-range neuronal communication, various experimental paradigms demanding attentive processing have shown long-range synchronization in a broad beta band, although mostly at frequencies below 25 Hz. The following provide a few examples of beta-band modulation in recent studies using very different task paradigms: Variations in reaction times and readiness to respond to a sensory-change event induced corresponding fine-grained variations of motor-spinal coherence in the beta band (Schoffelen et al., 2005). Somatosensory and motor cortex synchronize in the beta band during sensorimotor integration (Brovelli et al., 2004). Selective working memory maintenance in a delayed match-to-sample task results in stronger coherence in the beta band between higher visual areas in humans (Tallon-Baudry, Bertrand, & Fischer, 2001) and locally predicts performance in a similar task in the monkey (Tallon-Baudry, Mandon, Freiwald, & Kreiter, 2004). The failure to detect a target stimulus in a rapid stream of stimuli in the attentional blink paradigm is associated with reduced frontoparietal and frontotemporal beta-band synchronization (Gross et al., 2004). And as a last example for a potential functional role of beta-band activity, the perception of coherent objects from fragmented visual scenes goes along with transiently enhanced beta-band synchronization of the LFP among prefrontal, hippocampal, and lateral occipital sites (Sehatpour et al., 2008).

Taken together, these diverse findings agree to suggest that inter-areal synchronization critically subserves neuronal interactions during attentive processing. In the surveyed studies, synchronization in a broadly defined beta band occurred selectively during task epochs requiring effective neuronal integration of information across distributed cortical areas. However, further studies need to elucidate the properties of particular frequency bands and their characteristic recruitment during specific tasks (Kopell, Ermentrout, Whittington, & Traub, 2000).

Concluding remarks

Selective attention describes a central top-down process that restructures neuronal activity patterns to establish a selective representation of behavioral relevance. The surveyed evidence suggests that attention achieves this functional role by selectively synchronizing those neuronal groups conveying task-relevant information. Attentionally modulated synchronization patterns evolve rapidly, are evident even before sensory inputs arrive, follow closely subjective readiness to process information in time, can be sustained for prolonged time periods, and carry specific information about top-down selected sensory features and motor aspects.

In addition to these functional characteristics, insights into the physiological origins of synchronization have begun to shed light on the mechanistic underpinning of selective neuronal interaction patterns at all spatial scales of cortical processing: At the level of single neurons and local microcircuits, studies are deciphering the role of inhibitory interneuron networks, how precise timing information is conveyed and sustained even at high oscillation frequencies, and how rhythmic synchronization among interneurons is actively made robust against external influences (Bartos et al., 2007; Vida et al., 2006). These insights are integrated at the network level in models demonstrating how selective synchronization patterns evolve in a self-organized way (Börgers & Kopell, 2008; Tiesinga et al., 2008).

Acknowledging those basic physiological processes underlying the dynamic generation of selective synchronization seems to be pivotal to further elucidation of the mechanistic working principles of selective attention in the brain.

ACKNOWLEDGMENTS This work was supported by the Human Frontier Science Program Organization, the Volkswagen Foundation, the European Science Foundation's European Young Investigator Award program (PF), and the Netherlands Organization for Scientific Research (PF and TW).

REFERENCES

Azouz, R. (2005). Dynamic spatiotemporal synaptic integration in cortical neurons: Neuronal gain, revisited. *J. Neurophysiol.*, *94*(4), 2785–2796.

Azouz, R., & Gray, C. M. (2003). Adaptive coincidence detection and dynamic gain control in visual cortical neurons in vivo. *Neuron*, *37*(3), 513–523.

Bartos, M., Vida, I., & Jonas, P. (2007). Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nat. Rev. Neurosci.*, *8*(1), 45–56.

Bauer, M., Oostenveld, R., Peeters, M., & Fries, P. (2006). Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. *J. Neurosci.*, *26*(2), 490–501.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*(5721), 529–534.

Börgers, C., Epstein, S., & Kopell, N. J. (2005). Background gamma rhythmicity and attention in cortical local circuits: A computational study. *Proc. Natl. Acad. Sci. USA*, *102*(19), 7002–7007.

Börgers, C., & Kopell, N. (2003). Synchronization in networks of excitatory and inhibitory neurons with sparse, random connectivity. *Neural Comput.*, *15*(3), 509–538.

Börgers, C., & Kopell, N. J. (2008). Gamma oscillations and stimulus selection. *Neural Comput.*, *20*(2), 383–414.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., & Bressler, S. L. (2004). Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. *Proc. Natl. Acad. Sci. USA*, *101*(26), 9849–9854.

Buia, C., & Tiesinga, P. (2006). Attentional modulation of firing rate and synchrony in a model cortical network. *J. Comput. Neurosci*, *20*(3), 247–264.

Buia, C. I., & Tiesinga, P. H. (2008). The role of interneuron diversity in the cortical microcircuit for attention. *J. Neurophysiol.*, *99*(5), 2158–2182.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*(5820), 1860–1862.

Buzsaki, G. (2006). *Rhythms of the brain*. Oxford, UK: Oxford University Press.

Buzsaki, G., Kaila, K., & Raichle, M. (2007). Inhibition and brain work. *Neuron*, *56*(5), 771–783.

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., et al. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, *313*(5793), 1626–1628.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nat. Neurosci.*, *7*(3), 308–313.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345–347.

Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O., & Somogyi, P. (1995). Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature*, *378*(6552), 75–78.

Csicsvari, J., Jamieson, B., Wise, K. D., & Buzsaki, G. (2003). Mechanisms of gamma oscillations in the hippocampus of the behaving rat. *Neuron*, *37*(2), 311–322.

Doesburg, S. M., Roggeveen, A. B., Kitajo, K., & Ward, L. M. (2008). Large-scale gamma-band phase synchronization and selective attention. *Cereb. Cortex*, *18*(2), 386–396.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.*, *2*(10), 704–716.

Engel, A. K., Konig, P., Gray, C. M., & Singer, W. (1990). Stimulus-dependent neuronal oscillations in cat visual cortex: Inter-columnar interaction as determined by cross-correlation analysis. *Eur. J. Neurosci.*, *2*(7), 588–606.

Fan, J., Byrne, J., Worden, M. S., Guise, K. G., McCandliss, B. D., Fossella, J., et al. (2007). The relation of brain oscillations to attentional networks. *J. Neurosci.*, *27*(23), 6197–6206.

Frien, A., Eckhorn, R., Bauer, R., Woelbern, T., & Gabriel, A. (2000). Fast oscillations display sharper orientation tuning than slower components of the same recordings in striate cortex of the awake monkey. *Eur. J. Neurosci.*, *12*(4), 1453–1465.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn. Sci.*, *9*(10), 474–480.

Fries, P., Nikolic, D., & Singer, W. (2007). The gamma cycle. *Trends Neurosci.*, *30*(7), 309–316.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchroni-zation by selective visual attention. *Science*, *291*(5508), 1560–1563.

Fries, P., Schröder, J. H., Roelfsema, P. R., Singer, W., & Engel, A. K. (2002). Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *J. Neurosci.*, *22*(9), 3739–3754.

Fries, P., Womelsdorf, T., Oostenveld, R., & Desimone, R. (2008). The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4. *J. Neurosci.*, *28*(18), 4823–4835.

Ghose, G. M., & Maunsell, J. H. (2002). Attentional modulation in visual cortex depends on task timing. *Nature*, *419*(6907), 616–620.

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*(5), 677–696.

Gottlieb, J. (2002). Parietal mechanisms of target representation. *Curr. Opin. Neurobiol.*, *12*(2), 134–140.

Gray, C. M., Engel, A. K., Konig, P., & Singer, W. (1990). Stimulus-dependent neuronal oscillations in cat visual cortex: Receptive field properties and feature dependence. *Eur. J. Neurosci.*, *2*(7), 607–619.

Gray, C. M., Konig, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, *338*(6213), 334–337.

Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., et al. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proc. Natl. Acad. Sci. USA*, *101*(35), 13050–13055.

Gross, J., Schnitzler, A., Timmermann, L., & Ploner, M. (2007). Gamma oscillations in human primary somatosensory cortex reflect pain perception. *PLoS Biol.*, *5*(5), e133.

Gruber, T., Müller, M. M., Keil, A., & Elbert, T. (1999). Selective visual-spatial attention alters induced gamma band responses in the human EEG. *Clin. Neurophysiol.*, *110*(12), 2074–2085.

Hasenstaub, A., Shu, Y., Haider, B., Kraushaar, U., Duque, A., & McCormick, D. A. (2005). Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron*, *47*(3), 423–435.

Hauck, M., Lorenz, J., & Engel, A. K. (2007). Attention to painful stimulation enhances gamma-band activity and synchronization in human sensorimotor cortex. *J. Neurosci.*, *27*(35), 9270–9277.

Herculano-Houzel, S., Munk, M. H., Neuenschwander, S., & Singer, W. (1999). Precisely synchronized oscillatory firing patterns require electroencephalographic activation. *J. Neurosci.*, *19*(10), 3992–4010.

Hoogenboom, N., Schoffelen, J. M., Oostenveld, R., Parkes, L. M., & Fries, P. (2006). Localizing human visual gamma-band activity in frequency, time and space. *NeuroImage.*, *29*(3), 764–773.

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nat. Neurosci.*, *8*(2), 234–241.

Jin, Y., O'Halloran, J. P., Plon, L., Sandman, C. A., & Potkin, S. G. (2006). Alpha EEG predicts visual reaction time. *Int. J. Neurosci.*, *116*(9), 1035–1044.

Jones, M. W., & Wilson, M. A. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS Biol.*, *3*(12), e402.

Kaiser, J., Hertrich, I., Ackermann, H., & Lutzenberger, W. (2006). Gamma-band activity over early sensory areas predicts detection of changes in audiovisual speech stimuli. *NeuroImage*, *30*(4), 1376–1382.

Kayser, C., & König, P. (2004). Stimulus locking and feature selectivity prevail in complementary frequency ranges of V1 local field potentials. *Eur. J. Neurosci.*, *19*(2), 485–489.

Kelly, S. P., Lalor, E. C., Reilly, R. B., & Foxe, J. J. (2006). Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. *J. Neurophysiol.*, *95*(6), 3844–3851.

Khayat, P. S., Spekreijse, H., & Roelfsema, P. R. (2006). Attention lights up new object representations before the old ones fade away. *J. Neurosci.*, *26*(1), 138–142.

Kopell, N., Ermentrout, G. B., Whittington, M. A., & Traub, R. D. (2000). Gamma rhythms and beta rhythms have different synchronization properties. *Proc. Natl. Acad. Sci. USA*, *97*(4), 1867–1872.

Kreiter, A. K., & Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *J. Neurosci.*, *16*(7), 2381–2396.

Lakatos, P., Karmos, G., Metha, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*(5872), 110–113.

Landau, A. N., Esterman, M., Robertson, L. C., Bentin, S., & Prinzmetal, W. (2007). Different effects of voluntary and involuntary attention on EEG activity in the gamma band. *J. Neurosci.*, *27*(44), 11986–11990.

Lee, H., Simpson, G. V., Logothetis, N. K., & Rainer, G. (2005). Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron*, *45*(1), 147–156.

Lin, S. C., Gervasoni, D., & Nicolelis, M. A. (2006). Fast modulation of prefrontal cortex activity by basal forebrain noncholinergic neuronal ensembles. *J. Neurophysiol.*, *96*(6), 3209–3219.

Liu, J., & Newsome, W. T. (2006). Local field potential in cortical area MT: Stimulus tuning and behavioral correlations. *J. Neurosci.*, *26*(30), 7779–7790.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.*, *77*(1), 24–42.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.*, *5*(10), 793–807.

Markram, H., Wang, Y., & Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. USA*, *95*(9), 5323–5328.

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.*, *14*(9), 744–751.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends Neurosci.*, *29*(6), 317–322.

Mishra, J., Fellous, J. M., & Sejnowski, T. J. (2006). Selective attention through phase relationship of excitatory and inhibitory input synchrony in a model cortical neuron. *Neural Net.*, *19*(9), 1329–1346.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron*, *55*(1), 131–141.

Monosov, I. E., Trageser, J. C., & Thompson, K. G. (2008). Measurements of simultaneously recorded spiking activity and local field potentials suggest that spatial selection emerges in the frontal eye field. *Neuron*, *57*(4), 614–625.

Montgomery, S. M., & Buzsaki, G. (2007). Gamma oscillations dynamically couple hippocampal CA3 and CA1 regions during memory task performance. *Proc. Natl. Acad. Sci. USA*, *104*(36), 14495–14500.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784.

Munk, M. H., Roelfsema, P. R., Konig, P., Engel, A. K., & Singer, W. (1996). Role of reticular activation in the modulation of intracortical synchronization. *Science*, *272*(5259), 271–274.

Ohara, S., Crone, N. E., Weiss, N., & Lenz, F. A. (2006). Analysis of synchrony demonstrates "pain networks" defined by rapidly switching, task-specific, functional connectivity between pain-related cortical structures. *Pain*, *123*(3), 244–253.

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., & Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.*, *5*(8), 805–811.

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annu. Rev. Neurosci.*, *27*, 611–647.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.*, *19*(5), 1736–1753.

Riehle, A. (2005). Preparation for action: One of the key functions of motor cortex. In A. Riehle & E. Vaadia (Eds.), *Motor cortex in voluntary movements: A distributed system for distributed functions* (Vol. 1, pp. 213–240). Boca Raton, FL: CDC Press.

Rihs, T. A., Michel, C. M., & Thut, G. (2007). Mechanisms of selective inhibition in visual spatial attention are indexed by alpha-band EEG synchronization. *Eur. J. Neurosci.*, *25*(2), 603–610.

Rodriguez, R., Kallenbach, U., Singer, W., & Munk, M. H. (2004). Short- and long-term effects of cholinergic modulation on gamma oscillations and response synchronization in the visual cortex. *J. Neurosci.*, *24*(46), 10369–10378.

Roelfsema, P. R., Engel, A. K., König, P., & Singer, W. (1997). Visuomotor integration is associated with zero timelag synchronization among cortical areas. *Nature*, *385*(6612), 157–161.

Roelfsema, P. R., Tolboom, M., & Khayat, P. S. (2007). Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron*, *56*(5), 785–792.

Rudolph, M., Pospischil, M., Timofeev, I., & Destexhe, A. (2007). Inhibition determines membrane potential dynamics and controls action potential generation in awake and sleeping cat cortex. *J. Neurosci.*, *27*(20), 5280–5290.

Saalmann, Y. B., Pigarev, I. N., & Vidyasagar, T. R. (2007). Neural mechanisms of visual attention: How top-down feedback highlights relevant locations. *Science*, *316*(5831), 1612–1615.

Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.*, *2*(8), 539–550.

Sauseng, P., Klimesch, W., Freunberger, R., Pecherstorfer, T., Hanslmayr, S., & Doppelmayr, M. (2006). Relevance of EEG

alpha and theta oscillations during task switching. *Exp. Brain Res.*, *170*(3), 295–301.

Scherberger, H., & Andersen, R. A. (2007). Target selection signals for arm reaching in the posterior parietal cortex. *J. Neurosci.*, *27*(8), 2001–2012.

Scherberger, H., Jarvis, M. R., & Andersen, R. A. (2005). Cortical local field potential encodes movement intentions in the posterior parietal cortex. *Neuron*, *46*(2), 347–354.

Schoffelen, J. M., Oostenveld, R., & Fries, P. (2005). Neuronal coherence as a mechanism of effective corticospinal interaction. *Science*, *308*(5718), 111–113.

Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy effect in episodic memory. *NeuroImage*, *32*(3), 1422–1431.

Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J., & Madsen, J. R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *J. Neurosci.*, *23*(34), 10809–10814.

Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., et al. (2007). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cereb. Cortex*, *17*(5), 1190–1196.

Sehatpour, P., Molholm, S., Schwartz, T. H., Mahoney, J. R., Mehta, A. D., Javitt, D. C., et al. (2008). A human intracranial study of long-range oscillatory coherence across a frontal-occipital-hippocampal brain network during visual object processing. *Proc. Natl. Acad. Sci. USA*, *105*(11), 4399–4404.

Sejnowski, T. J., & Paulsen, O. (2006). Network oscillations: Emerging computational principles. *J. Neurosci.*, *26*(6), 1673–1676.

Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *J. Neurosci.*, *21*(4), 1340–1350.

Siegel, M., & König, P. (2003). A functional gamma-band defined by stimulus-dependent synchronization in area 18 of awake behaving cats. *J. Neurosci.*, *23*(10), 4251–4260.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends Cogn. Sci.*, *9*(1), 16–20.

Tallon-Baudry, C., Bertrand, O., & Fischer, C. (2001). Oscillatory synchrony between human extrastriate areas during visual short-term memory maintenance. *J. Neurosci.*, *21*(20), RC177.

Tallon-Baudry, C., Mandon, S., Freiwald, W. A., & Kreiter, A. K. (2004). Oscillatory synchrony in the monkey temporal lobe correlates with performance in a visual short-term memory task. *Cereb. Cortex*, *14*(7), 713–720.

Taylor, K., Mandon, S., Freiwald, W. A., & Kreiter, A. K. (2005). Coherent oscillatory activity in monkey area v4 predicts successful allocation of attention. *Cereb. Cortex*, *15*(9), 1424–1437.

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.*, *26*(37), 9494–9502.

Tiesinga, P., Fellous, J. M., & Sejnowski, T. J. (2008). Regulation of spike timing in visual cortical circuits. *Nat. Rev. Neurosci.*, *9*(2), 97–107.

Tiesinga, P. H., Fellous, J. M., Salinas, E., Jose, J. V., & Sejnowski, T. J. (2004). Inhibitory synchrony as a mechanism for attentional gain modulation. *J. Physiol. Paris*, *98*(4–6), 296–314.

Tiesinga, P. H., & Sejnowski, T. J. (2004). Rapid temporal modulation of synchrony by competition in cortical interneuron networks. *Neural Comput.*, *16*(2), 251–275.

Tiitinen, H., Sinkkonen, J., Reinikainen, K., Alho, K., Lavikainen, J., & Naatanen, R. (1993). Selective attention enhances the auditory 40-Hz transient response in humans. *Nature*, *364*(6432), 59–60.

Treue, S., & Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579.

Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nat. Rev. Neurosci.*, *2*(4), 229–239.

Vida, I., Bartos, M., & Jonas, P. (2006). Shunting inhibition improves robustness of gamma oscillations in hippocampal interneuron networks by homogenizing firing rates. *Neuron*, *49*(1), 107–117.

von Stein, A., Chiang, C., & König, P. (2000). Top-down processing mediated by interareal synchronization. *Proc. Natl. Acad. Sci. USA*, *97*(26), 14748–14753.

Wannig, A., Rodriguez, V., & Freiwald, W. A. (2007). Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron*, *54*(4), 639–651.

Womelsdorf, T., & Fries, P. (2007). The role of neuronal synchronization in selective attention. *Curr. Opin. Neurobiol.*, *17*(2), 154–160.

Womelsdorf, T., Fries, P., Mitra, P. P., & Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature*, *439*(7077), 733–736.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., et al. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science*, *316*(5831), 1609–1612.

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *J. Neurosci.*, *20*(6), RC63.

Wrobel, A., Ghazaryan, A., Bekisz, M., Bogdan, W., & Kaminski, J. (2007). Two streams of attention-dependent beta activity in the striate recipient zone of cat's lateral posterior-pulvinar complex. *J. Neurosci.*, *27*(9), 2230–2240.

Wyart, V., & Tallon-Baudry, C. (2008). Neural dissociation between visual awareness and spatial attention. *J. Neurosci.*, *28*(10), 2667–2679.

Yamagishi, N., Callan, D. E., Goda, N., Anderson, S. J., Yoshida, Y., & Kawato, M. (2003). Attentional modulation of oscillatory activity in human visual cortex. *NeuroImage*, *20*(1), 98–113.

IV

SENSATION AND PERCEPTION

Chapter 21 BARLOW 309

22 YESHURUN, LAPID, HADDAD, GELSTIEN, ARZI, SELA, WEISBROD, KHAN, AND SOBEL 321

23 RICHARDS AND KIDD 343

24 WRIGHT AND ZHANG 353

25 GRIFFITHS, KUMAR, VON KRIEGSTEIN, OVERATH, STEPHAN, FRISTON 367

26 CARROLL, YOON, AND WILLIAMS 383

27 BRAINARD 395

28 RINGACH 409

29 SEIDEMANN, CHEN, AND GEISLER 419

30 GOEBEL AND DE WEERD 435

31 CONNOR, PASUPATHY, BRINCAT, AND YAMANE 455

32 McKONE, CROOKES, AND KANWISHER 467

33 DeANGELIS 483

34 ANGELAKI, GU, AND DeANGELIS 499

35 MORRONE AND BURR 511

36 SIMONCELLI 525

Introduction

J. ANTHONY MOVSHON AND BRIAN A. WANDELL

THIS IS THE FOURTH edition of *The Cognitive Neurosciences*. With each succeeding edition comes the same question: Why does a volume on cognition include sensation and perception—and a double dose at that? The answers to these questions lie in the chapters of this section.

For more than a century, there has been a systematic quest to understand what and how physical information is captured by the nervous system. A great gift of 20th-century neuroscience is a set of beautiful and definitive experiments that explain much about this peripheral encoding. Through the use of illusions and demonstrations, introductory textbooks rightly emphasize that these peripheral signals must be interpreted for their biological significance. For example, the signals encoded in the retina do not measure the distance to a predator; the signals encoded in the auditory nerve do not specify the significance of a sound. Rather, as Helmholtz taught us, the brain uses the receptor signals to make inferences about information. The ongoing work in sensation and perception has shifted to focus on questions about the brain's computational circuitry: What are the computations performed on raw sense data by the sensory systems? What quantities of interest are extracted by these computations? What neural circuits perform them, and how do these circuits work?

Science is frequently imprisoned by technology, and neuroscience is no exception. Our view of brain function has been conditioned and distorted by the methods that we have available. In sensory neuroscience, the dominant technique for most of the last quarter-century has been single-unit recording. Unit recording experiments offer unparalleled access to the details of neural computation and allow us to measure activity at a fundamental computational scale of the system. But the resolution of unit recording comes at a cost. Brain function depends on the organization and trans-

mission of information over distances far larger than a single electrode can sense, and we have had a difficult time obtaining a view of the nervous system's actions at long spatial scales. Our view is expanding on the basis of new methods for measuring neural signals at multiple length scales. There are noninvasive methods to measure the spatial organization of cone photoreceptors in the living human eye and new methods to measure spatially resolved activity in the human brain. Techniques for studying cells now measure dozens of cells at once and can even track important activation-related signals from local clusters of thousands of neurons. These advances all provide valuable information about the way in which sensory signals are encoded by populations of neurons and how these signals are interpreted within the brain. The new measurements are accompanied by developments in computational principles and tools that help us to understand how the nervous system extracts biologically relevant events from a complex array of environmental signals. The chapters in this section show how the combination of new measurements and analytic methods are being combined to create a deeper understanding of how the nervous system represents biologically relevant sensory events.

The topics in this section span vision, audition, olfaction, analysis, and theory. The reader will find a wide array of techniques, including anatomy, electrophysiology, optical imaging, neuroimaging, behavior, and computation. The close connection between computation and methods is nicely illustrated in the chapters concerned with peripheral encoding. Brainard considers how one might make unexpected inferences about object color from the encoding by the three types of cones; Carroll, Yoon, and Williams analyze the information contained in the cone spatial mosaic. The computational analysis of neural encoding is fundamental across sensation. Yeshurun and colleagues describe how psychophysical and computation methods are used to understand the organization of olfaction. Richards and Kidd introduce the reader to questions about how the auditory system disentangles sounds in a complex environment. Two theoretical chapters, by Simoncelli and Barlow, consider approaches to analyzing information in sensory arrays. Simoncelli offers a modern approach that builds on the fundamental work of important pioneers; Barlow, one of these pioneers, gives us a historical perspective and puts his current thinking in that context.

Much important work remains to be done at the single-neuron level, and each of the chapters by Connor and colleagues, DeAngelis, and Angelaki and colleagues offers a snapshot of the state of the art. Connor and colleagues show how the rigorous application of quantitative analysis methods can begin to offer an account of the transformation of information about visual sensory elements in V1 into a more subtle and sophisticated representation of object features in temporal cortex. DeAngelis considers area MT, best known for its contributions to visual motion perception, and gives an account that combines unit recording and behavioral analysis to demonstrate that MT also plays a key role in the perception of depth. Angelaki and colleagues remind us that visual signals are processed in combination with information from other senses, and describes a series of experiments showing the coordinated role that vestibular and visual signals play in representing information about motion through the world in areas that used to be considered "purely" visual. There are intimate links between sensory and motor systems, none closer than the link between the visual system and the oculomotor system. Morrone and Burr review behavioral and functional imaging work, addressing the ways in which these systems work together to compute our stable perceptual experience from a highly unstable retinal signal.

Many researchers now seek to take data from the unit level and generalize it into a form that can be used to account directly for behavior. Ringach reexamines the specificity of neuronal response in the most-studied of all sensory areas—V1—and shows unexpected relationships between that specificity and the cortex's functional architecture. Seidemann and colleagues also start with the familiar visual representation in V1 but use optical imaging techniques to ask about population activity on a scale of millimeters rather than microns. Drawing on his recordings, on simultaneous behavioral measurement, on knowledge from unit work, and on the theory of encoding and representation, he constructs an account of simple visual performance.

The revolution in functional MRI, enabling scientists to make spatially resolved measurements of the awake-behaving human brain, has also played an important role in sensation and perception. The chapters by McKone, Griffiths, and Goebel and their colleagues describe approaches to interpreting signals in human visual cortex. Taking a developmental approach, McKone and colleagues seek to understand how the pathways that are needed to recognize and interpret faces develop. Griffiths and colleagues apply computational methods to understanding the structure of sounds and then further considers ideas about how the responses spread across cortex might interpret these sounds. Goebel and De Weerd examine the process of image interpolation (filling-in) in visual cortex. Notions of interpretation are often connected with neural plasticity (see Section II), which also plays a major role in sensory processing. Wright and Zhang document a series of behavioral studies showing how auditory information processing depends on training and experience, suggesting that much of the cortical machinery studied by others can adapt its function to suit the needs of the organism.

Taken together, these contributions show that sensation and perception are intertwined with cognition in two fundamental ways. First, the work in sensation and perception

crosses borders between many fields, integrating work in behavior, neuroscience, and computation. This field offers an excellent testing ground to evaluate many of the techniques that will be needed as we develop the field of cognitive neuroscience. Second, the chapters in this section show clearly that the information that sensory and perceptual pathways provide to cognition sets critical bounds on the information that is available to the brain. These are limits that influence the thoughts we have, the decisions we make, and the emotions we experience. Perception provides three dimensions of color and three dimensions of space; there are molecules whose smell evokes pleasure or disgust. Sensation and perception serve the needs of cognition; cognition, in turn, works within the bounds set by perception.

21 Grandmother Cells, Symmetry, and Invariance: How the Term Arose and What the Facts Suggest

HORACE BARLOW

ABSTRACT By the late 1960s, recording from sensory pathways had shown that single neurons can be much more sensitive, selective, and reliable than had previously been recognized. The term *grandmother cell* started as a fanciful name for a high-level neuron that might enable us to experience complex perceptions and discriminate among them. The concept included invariance of response for changes in some variables as well as selectivity of response for others, together with the idea that these cells are created by processing at a hierarchy of levels. This chapter first outlines the discoveries that eventually led to the general acceptance that such cells really exist. It then discusses hierarchical processing, the evolution of the cortex, and ideas about the new behavioral faculties that evolved with it. Finally it points to the enormous, unaccounted for number of neurons in the cortex and suggests that this plays a major role in enhancing our ability to exploit symmetry and invariance in our environment.

The term *grandmother cell* suggests that there are particular neurons in the visual cortex that are activated by the sight of one's grandmother, and it implies that such neurons play an important part in generating high-level perceptions and discriminating between them. The term was introduced in 1969 by Jerry Lettvin (see Barlow 1995), and in the first part of this chapter, I shall present a brief history of the facts as they have been discovered since then. In the early years, many people thought that it was just a catchy term for an implausible idea that would lead nowhere, but by now, it is clear that neurons fitting this definition really exist and that advances in understanding their neurophysiology will constitute real progress in understanding the brain, particularly the conscious, thinking parts of it.

Even before Lettvin had coined the term, Konorski (1967) had championed the idea that what he called *gnostic neurons* were the end result of a hierarchical series of transformations along the lines that Hubel and Wiesel (1962, 1965) had suggested from the results of their single-unit recordings in the visual cortex. Although this idea of a hierarchy has not completely crashed, it has had a bumpy ride and never specified a functional goal that the supposed hierarchy might help to achieve.

The second part of this chapter briefly considers some objections to the idea of a hierarchy, but we rapidly encounter the fact that the cortex contains vastly more neurons representing each location in the visual field than the retina or LGN, and the need for this vast excess has not been explained by any current computational model. This makes one suspect that the cortex performs computations that are different from those of earlier stages in the visual pathway or other parts of the brain, which adds a new slant to the problem. To help decide what these new computations might be, in the third part I look at the problem from a broader perspective, consulting other academic disciplines, and this leads to the idea that the cortex uses the symmetry and invariance present in its input to generate a more economical, sparser, representation of the environment.

A brief history

Charles Gross deserves at least as much credit as anyone else for the actual experiments on which claims about grandmother cells rest, and in his essay "Genealogy of the Grandmother Cell" (Gross, 2002), he tells how Jerzy Konorski's neuropsychological studies of visual agnosia, with evidence of the functional hierarchy of neurons in the visual cortex from Hubel and Wiesel, further helped by Jerry Lettvin's fertile imagination, may have guided his laboratory toward these discoveries. Gross gave the definition shown in box 21.1, and this will be adopted as an initial working definition, since it coincides well with most people's usage. However, one of the main points of the grandmother cell idea is that the cells' responses are largely unaffected by changes of grandma's position, pose, clothing, facial expression, and so on. Invariance has got lost in box 21.1's definition, but it is at least as important a property as selectivity is.

HORACE BARLOW Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom

Box 21.1 Definition of Grandmother Cells by Charles Gross (2002)

A "grandmother cell" is a hypothetical neuron that responds only to a highly complex, specific, and meaningful stimulus, such as the image of one's grandmother.

Konorski (1903–1973) was a psychologist who had worked with Pavlov and is best known for making the distinction between what are now termed *operant conditioning* and *Pavlovian conditioning*. He was also a clinical neurologist and was particularly interested in disturbances in the understanding and production of speech. In some patients, the disturbance is confined to a specific category of objects; for instance, one patient might confuse the names of small human-made objects (figure 21.1, line a), with little or no evidence of confusions within categories sketched on other lines, whereas another patient's problems might be confined to animate objects (figure 21.1, line f). Konorski suggested this is because the relevant part of the cortex is subdivided spatially into distinct regions that he called *gnostic fields*, so local damage could affect one category without affecting others. For future reference, note the bottom line (figure 21.1, line i), suggesting there is a gnostic field that specializes in body parts.

Konorski was aware of the work of Hubel and Wiesel (1962, 1965) on the selective sensitivity of individual neurons in the visual cortex, and when they suggested that these neurons are arranged in a functional hierarchy, he took this up and proposed that his gnostic fields are populated by *gnostic neurons*, whose job it is to respond to images of particular objects within the semiological category for the gnostic field they are located in. For example, in the "small human-made objects" gnostic field, he thought that some neurons might respond when a key was shown to the patient and others when a pair of spectacles was shown and that these different responses would enable them to be distinguished and appropriately named by the patient.

Gross reviewed *The Integrative Action of the Brain (*Konorski, 1967) and was mightily impressed by it, so he started testing Konorski's ideas in his lab, where he was recording from neurons in monkey inferotemporal cortex. Gross's lab was then at the Massachusetts Institute of Technology, where Jerry Lettvin daily preached his inspired views in his own imaginative style. I was once accused of inventing grandmother cells and wrote to Jerry Lettvin to ask whether he was the real hero (or villain). He immediately confessed, explaining that in 1969, he had had to take over a lecture course at short notice and wanted to illustrate how a neuron might represent a rather complex concept. He therefore prepared the story sketched in box 21.2.

Box 21.2 Lettvin's Story Illustrating the Grandmother Cell Concept

In the distant Ural mountains lives my second cousin, Akakhi Akakievitch, a great if unknown neurosurgeon. Convinced that ideas are contained in specific cells, he had decided to find those concerned with a most primitive and ubiquitous substance—mother. Starting with geese and bears . . . he progressed to Trotskyites under death sentence in Siberia. . . . He located some 18,000 neurons clustered in von Seelendonck's area . . . that responded uniquely only to the animal's mother, however displayed . . . animate or stuffed . . . offered by caricature, photograph, or abstraction . . . etc., etc.

The story went on for many more paragraphs and ended up with poor Akakhi, on the threshold of a Nobel Prize, having to start all over again with grandmother cells because, for reasons that are beyond me, mother cells with such properties would be politically unacceptable. (With his permission, Lettvin's letter was reproduced in full in an appendix in Barlow, 1995.) Note particularly the last two lines of Lettvin's story, emphasizing the invariance of the selectivity of the postulated neurons for many types of change in the stimulus.

It must surely have been the experimental results obtained by recording the activities of neurons in sensory pathways that gave Jerry Lettvin the idea that a complex concept could be represented by the firing of a single neuron, for by the late 1960s, this work had shown that single neurons were much more sensitive, reliable, and selective in their responses than had previously been supposed. The facts and arguments that led to this conclusion can be found in my review (Barlow, 1972), but three points about this early work deserve attention.

First, a lot of the early research came from outside the field of vertebrate neurophysiology, for example, work on insects and crustaceans (Wiersma, Waterman, & Bush, 1961; Hartline, 1949; Reichardt, 1961). Second, the field was fertilized by ideas from different, apparently unrelated, fields; I was myself much influenced by the idea of innate releasing factors and fixed action patterns from ethology and by the importance of signal-to-noise ratios from signal detection theory and feature detectors from early work on pattern recognition and computer vision. Third, the work of Lettvin and his colleagues on the frog visual system (Lettvin, Maturana, McCulloch, & Pitts, 1959), though not reported with enough detail to satisfy some of their colleagues, was certainly effective in spreading the important message that single neurons respond reliably, and in some cases invariantly, to complex and meaningful features in the visual field.

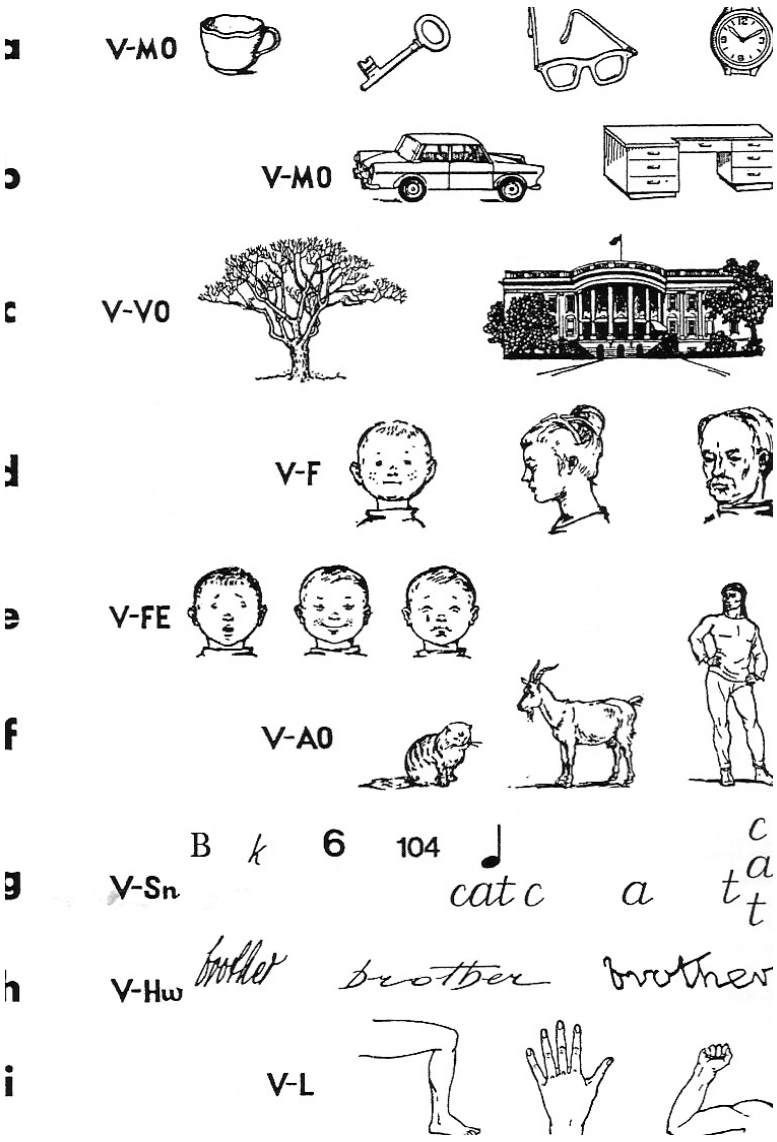


FIGURE 21.1 Konorski's gnostic fields and neurons. Clinical evidence suggests that aphasia is sometimes partial, affecting the ability to name or recognize objects in specific semiological categories, such as small human-made objects (line a), facial expression (line e), or body parts (line i). When Hubel and Wiesel (1962, 1965) put forward the hierarchical view of cortical processing, Konorski suggested that single gnostic neurons within each gnostic field responded selectively to different items within the semiological category of that gnostic field. This is very similar to Lettvin's proposal (see box 21.2) but does not emphasize the invariance of their responses. (From Konorski, 1967; Gross, 2002.)

MIT lies across the Charles River from Hubel and Wiesel's lab at Harvard Medical School, from which the evidence for a hierarchical functional organization in the cortex had emerged. Hubel and Wiesel (1959) had discovered that neurons in V1 are selective for the orientation of visual stimuli, and I shall make a digression here to illustrate how difficult it was to make that step. This is based partly on my own experience somewhat later, when I was working with Bill Levick on retinal ganglion cells in the rabbit (Barlow, Hill, & Levick, 1964); these fall into many different categories whose high degree of selectivity for different, specific patterns makes them, in some respects, like precortical grandmother cells.

Most people who have recorded from neurons in sensory pathways will have experienced the long periods of intense frustration that occur when you know that your electrode is near a cell, because you detect its action potentials when it fires spontaneously, but you are unable to find the visual stimulus, nicknamed its "trigger-feature," that reliably excites it. In such cases, one frequently relieves one's frustration by moving on to another neuron in the hope that one will find its trigger feature more easily, but in the following quotation,

David Hubel describes how he and Torsten Wiesel persisted and were rewarded.

They had initially assumed, like others before them, that a uniform field of light would stimulate most cells in visual cortex, but they did not find any significant responses to such general changes of illumination. They therefore started projecting small dark or light spots onto a screen, because these are very effective stimuli for retinal ganglion cells and LGN neurons, so they at least knew that the inputs to the cortical neurons would be strongly active. What happened next is described by David Hubel in box 21.3.

Gross and his colleagues, exploring a new part of the cortex, were in an even more unpromising position, and it might have been in some desperation that they resorted to testing responses to objects for which Konorski had postulated "gnostic neurons." When they tried items for the "body parts" category (figure 21.1, line f), they obtained results such as those shown in figure 21.2, and many examples of neurons selective for the view of a monkey face were also found.

These amazing results initially evoked much skepticism, and many reasons for disbelieving them were advanced: Perhaps the view of another monkey's face or hand aroused the interest of the monkey whose cells were being probed, and this general excitement spread to the particular cell being recorded from. Without actually being present and observing many responses to many different stimuli, one can certainly be forgiven for being skeptical and cautious, for there is an almost infinite range of stimuli that could have been tested, and there is real difficulty in describing the range that was actually used and why these were selected. One wonders whether they ever heeded Lettvin's voice and tried a real grandmother!

Box 21.3 David Hubel and Torsten Wiesel Discover Orientational Selectivity (Hubel, 1988)

After about five hours of struggle, we suddenly had the impression that the glass with the dot was occasionally producing a response, but the response seemed to have nothing to do with the dot. Eventually we caught on: it was the sharp but faint shadow cast by the edge of the glass as we slid it into the slot that was doing the trick. We soon convinced ourselves that the edge worked only when its shadow was swept across one small part of the retina and that the sweeping had to be done with the edge in one particular orientation. . . .

The discovery was just the beginning. . . .

The neurophysiological world was slow to repeat these experiments, but more papers from Gross's own lab (Gross, Bender, & Rocha-Miranda, 1969; Gross, Rocha-Miranda, & Bender, 1972; Desimone, Albright, Gross, & Bruce, 1984), from Perrett, Rolls, and Caan (1982) in Oxford, and from Yamane, Kaji, and Kawano (1988) in Japan convinced most of us that neurons are to be found in monkey inferotemporal cortex (and elsewhere) that respond better to monkey-like faces than to other visual stimuli ("better" here means simply that more spikes were produced during the few hundred milliseconds following the stimulus). These responses can be produced from a wide range of positions in the visual field and over a wide range of distances, so they show considerable invariance for retinal image position and size while remaining selective for pattern. Different neurons varied a great deal in how selective they were for the pose or orientation of the stimulus, and they responded to both monkey and human faces. In some cases, it seemed clear that a good response was reserved for a particular human individual's face, among a dozen or so tested, regardless of pose, position, and distance. It also came to be appreciated that faces are peculiarly important for monkeys, and Perrett's work in particular (Perrett, Heitanen, Oram, & Benson, 1992) showed that some cells were responsive to aspects of the image presented that had special behavioral significance, such as the direction of a monkey's gaze.

One extraordinary development was the finding of face cells in humans (Kreiman, Koch, & Fried, 2000; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). Single neurons were recorded from the medial-temporal regions of patients whose brains were being explored in order to shed light on the origins of their epileptic attacks, and the researcher reported neurons that were responsive not only to images of the actresses Jennifer Aniston and Halle Berry, but also, in one instance, to the written name. It is amazing how this finding echoes the spirit of Lettvin's imaginary neurons that "responded uniquely only to the animal's mother, however displayed . . . animate or stuffed . . . offered by caricature, photograph, or abstraction." One particularly wants to know how the experimenters selected the range of stimuli that were tested. How did they happen to have available the

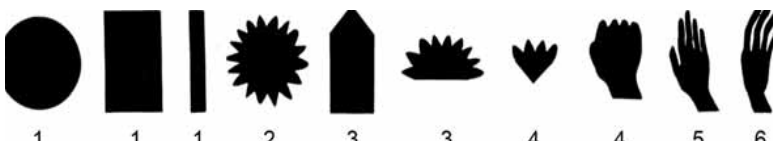


Figure 21.2 Examples of shapes used to stimulate a group TE unit apparently having a very complex trigger feature. The stimuli are arranged from left to right in order of increasing ability to drive the neuron from none (1) or little (2 and 3) to maximum (6). (From Gross, Rocha-Miranda, & Bender, 1972.)

picture of the right actress when testing a particular neuron? Although there are, inevitably, unanswered questions about such work, done under such very restrictive conditions, these results illustrate beautifully the property of selectivity for one characteristic of the stimulus (personal identity) while maintaining invariance for others (e.g. whether this identity is conveyed by a photograph or in writing).

Turning back to Gross's definition, we can see that he was perhaps overly modest, for by 2002, the existence of neurons that are very like "grandmother cells" was no longer hypothetical. We can also see that Gross was unfortunate in adopting the prototypical name Lettvin had chosen for explaining the concept of a class of cells that were selectively sensitive to "highly complex, specific, and meaningful stimuli"; no one has ever reported cells whose trigger feature was actually "grandmother." The term is flippant, incomplete, and not quite accurate, but it is widely understood, and I shall stick with it.

Although by the new millennium, the mere existence of grandmother cells could not seriously be doubted, their reported properties were very variable, and although they were known to be widely distributed, they were not thought to occur anywhere with high density. Furthermore, their functional roles and how they were wired up to produce their selective sensitivities and invariant responses remained unknown. But this was about to change.

The first important finding was that significant fMRI responses occur at several positions in scans of humans viewing whole faces but not fragmented faces or nonface objects (Puce, Allison, Gore, & McCarthy, 1995; Kanwisher, McDermott, & Chun, 1997). Apparently one of these face-sensitive patches, as they are called, the fusiform face area, even yields significant blood oxygen level dependent (BOLD) fMRI responses when the subject just thinks of faces (O'Craven & Kanwisher, 2000). Tsao, Freiwald, Knutsen, Mandeville, and Tootell (2003) showed that rhesus monkeys also have several face-selective patches, and detailed statistical analysis of the BOLD responses in different cortical positions to different stimulus objects showed that the pattern of activation produced by a stimulus object contained sufficient information to determine which of the objects was a face.

This is an intriguing finding that replicates similar analyses of human fMRI results (Haxby et al., 2001; Spiridon & Kanwisher, 2002). It shows that much information about the nature of a stimulus object can be derived from the distribution of activity in the cortex on the coarse spatial scale of voxels provided by the fMRI scans. It has also been shown that microstimulation in regions that are rich in face-selective neurons interferes with the correct categorization of noise-perturbed face images by awake, behaving monkeys (Afraz, Kiani, & Esteki, 2006). Since microstimulation is thought to act on a spatial scale considerably coarser than that available to single neurons, this result fits in with that of Haxby and colleagues (2001). However the fact that the object identification problem can be solved by using information segregated on a coarse scale and is interfered with by the coarse-scale interference of microstimulation by no means disproves the idea that the brain depends upon face-selective single neurons to recognize and discriminate between faces.

If the brain only uses the coarse-scale information revealed by fMRI, what part of the brain performs the complex statistical computation that Haxby and colleagues (2001) had to do on the fMRI data to get their result? But even though it can be interpreted in misleading ways, the fact that the coarse-scale localization of activity provides information for object identification remains important.

Kanwisher and colleagues (1997) also realized that fMRI scanning might open an opportunity for serious analysis of grandmother cells in monkeys, for it could perhaps tell neurophysiologists precisely where to place their electrodes to obtain a high yield of face-selective cells and thus make systematic examination of their varied properties possible, and perhaps it could also reveal the significance of there being several face-selective patches. Although such examination is still in its early stages, the technique is showing enormous promise.

In the first experiments in which they guided their recording electrode not just to the general region where previous investigators had found face cells, but to the region in a specific monkey where fMRI had revealed face-specific activity, Tsao and colleagues found 97% of neurons (out of about 400 recorded from) that showed greater response to pictures of whole faces than to nonface objects or to fragmented face images (Tsao, Freiwald, Tootell, & Livingston, 2006; Tsao, 2006). This compares with a figure no higher than 30% (and often much lower) for previous investigators who were not using fMRI guidance. Figure 21.3 shows three prominent face patches in the temporal lobe of a macaque, together with BOLD responses from a face patch as a function of time when stimulated by faces alternated with hands or other nonface objects. The third part of the figure compares the healthy responses to the faces with the insignificant or even weakly negative responses to all nonface objects, with the exception of an apple and a clock, which gave weak but significant positive responses that need no further comment.

It is interesting that the face stimuli that were used in these experiments were pictures of human faces, all presented in full-face, erect view. As might be expected, the monkeys' cortical neurons actually preferred monkey faces, consistently giving larger responses to them than to human faces. They also gave weak responses to a very crude face cartoon when it was erect but not when it was inverted, as shown in the bottom two lines in figure 21.3, and this was important, for it made possible a later, more detailed analysis of the

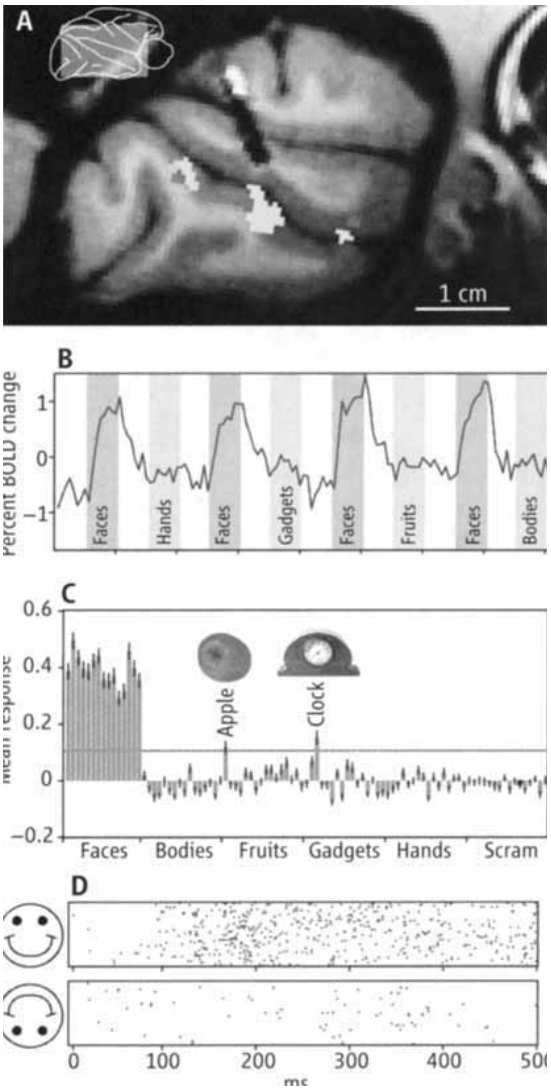


FIGURE 21.3 (*A*) Three patches of face-selective fMRI activation (yellow regions) in the macaque temporal lobe. (*B*) Time course from the face patches. Blood flows to these regions increase only when the monkey views faces. (*C*) Average response across 182 cells from the middle face patch of one monkey to 96 different images. The first 16 images are faces. (*D*) Responses of a face cell to repeated presentations of an upright and an inverted cartoon face. Each dot represents an action potential. (From Tsao, 2006.) (See color plate 21.)

basis for the pattern-selective properties of these neurons, using partial and complete face cartoons with graded feature properties. This part of their work is obviously based on a good knowledge of the face recognition literature, and the results are clearly described and thoroughly analyzed. The overall impression that is given is that different neurons respond to different snippets of information about faces, sometimes to single features such as iris size, sometimes to combinations of features. Whole faces usually give large responses, though these are never as large as the sum of the responses to individual features.

Many hypotheses about the nature of the selectivity were explored, and there is a lot of information against which one can test one's own individual theory of face recognition, but I think it is fair to say that clear models do not emerge, either for what the middle face patch and other face patches do or for how they may do it.

The good news, then, is that we have a system in which the pattern selectivity and invariance of face cells can be explored. The analysis of how neural mechanisms determine their selectivity for some features and invariance for others has not been solved, but it is on its way.

Cortex: Hierarchy, origin, and new behavior

Hubel and Wiesel's hierarchical proposal has been important for all our thinking about cortical function and is, for instance, a central feature both of the anatomical scheme of Felleman and van Essen (1991) and of the computational view of Riesenhuber and Poggio (1994), but it has also been criticized on a number of grounds. For example, hypercomplex cells are thought to be the same as what are now termed *end-stopped neurons*, and the details of the anatomical evidence do not coincide with the physiological evidence as nicely as Hubel and Wiesel would probably have liked; for instance, V1 has both simple and complex cells, which should not be the case, according to their scheme, if V1 is a single stage in the hierarchy. There is also good evidence that complex cells receive a direct input from the LGN, which again does not fit the suggested scheme exactly.

The agreed arrangement of the interconnected areas in the cortex is not obviously that of a hierarchy, for there is no "top area," and if one area connects directly to another, there is almost always a direct connection in the reverse direction as well. The term also introduces inappropriate implications from its everyday use to describe hierarchies of command in the army or of decision making in a commercial business, in which there are often very different numbers at different levels—one boss, many workers, for example—and the flow of command may be strictly unidirectional and may strictly avoid skipping levels. Also, the details often turn out to be different from those that were originally suggested; for example, there are orientationally selective ganglion cells in the retina (Levick, 1967), whereas this type of selectivity was initially not thought to occur before the primary visual cortex.

These may all be minor quibbles, but there is one fact suggesting that if there is a cortical hierarchy, it involves new methods or new principles not employed at earlier levels in the visual pathway. This is shown by doing counts of neurons at different levels in the visual pathway, for there is a large, abrupt rise on entering the cortex. Figure 21.4 shows the numbers of neurons in the precortical visual pathway and various parts of the cortex of the macaque. The main message stands out bold and clear: There is an enormous

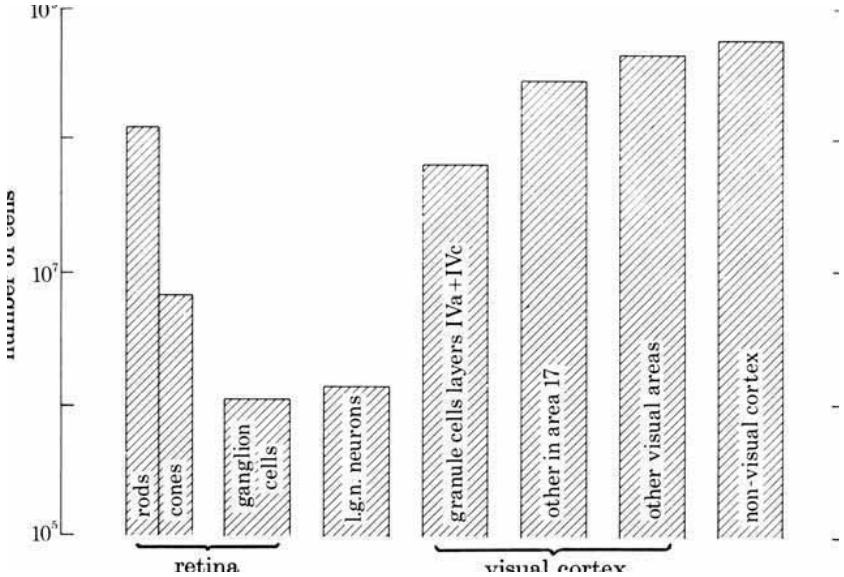


FIGURE 21.4 Estimated numbers of cells at various levels in the visual pathway of a macaque monkey. Note that the scale is logarithmic and that there are 55 times as many granule cells as LGN neurons. If one takes all the 250 million neurons in V1, this is 180 times the number of LGN neurons, and for the whole visual cortex (nearly 800 million neurons), it is almost 600 times. (From Barlow, 1981; adapted from Chow, Blum, & Blum, 1950.)

increase in the number of neurons involved in vision as the messages are passed from the LGN neurons to the granule cells in layer 4 of the primary visual cortex. The actual figures are interesting. Reading from the chart, this first stage of increase is by a factor of about 55, from 1.4 million to 75 million. If one includes all the 250 million neurons in V1, the factor increases to 180, and for the whole visual cortex (nearly 800 million neurons), the number of neurons is almost 600 times the number of LGN neurons.

That really establishes the point that the cortex has very many more cells at its disposal than do precortical levels, but the numbers in figure 21.4 represent averages over the whole visual field. It is estimated that in the foveal region, the density of neurons in V1 per unit solid angle of the visual field is 10,000 times the density of input fibers from the LGN (Hawken & Parker, 1991). This compares with the average over the whole visual field of 180 times, given above. The difference results from the fact that the cortical representation of the fovea has much more than its fair share of neurons, even after taking account of the overrepresentation of the fovea in the input from the LGN. The number of neurons available for computations on the input from the fovea is truly phenomenal.

Detailed accounting for the numbers shown in the left part of figure 21.4 is complicated by the fact that rods and cones behave differently, and the pattern of convergence and then divergence is very different at the fovea and in the periphery. Many of these matters are not relevant here, but there is one point that makes sense physically, and this is worth pointing out, lest the obvious message conveyed above be overshadowed by these complications.

The actual number of nerve fibers running in parallel from retina to LGN is not far from the number required for there to be one pathway per resolvable element of the optical image falling on the retina or one pathway per cone, provided that one confines one's attention to the foveal part of the pathway and takes into account that the image in the near periphery is undersampled (i.e., it has fewer nerve fibers serving it than the quality of the image deserves). The approximate agreement in the fovea means that with these cautions, the number of nerve fibers in the optic nerve roughly coincides with the number of degrees of freedom in the copy of the image that the optic nerve passes to the brain. Why, then, does the cortex need up to 10,000 times that number of neurons to perform its computations? The fact that we think we understand the limiting factors of foveal vision makes the huge numbers of neurons that are available in that part of the cortex even more impressive. Is there something about the computations it does that we have missed or paid insufficient attention to?

The next part of this chapter discusses this problem in light of the new behavior that the new part of the mammalian brain, the cerebral cortex, is thought to have brought about. Although this evidence is not very satisfying for a modern neurophysiologist, it is one of the few sources that can give useful hints about these new functions.

The cerebral cortex first appeared in the forebrain of mammals while dinosaurs were the dominant large, terrestrial vertebrates. Before mammals, the vertebrate forebrain had been dominated by its olfactory input, and in fact, olfaction is the only sensory modality that still has a direct input to the cortex; all the others pass through the thalamus. These

facts led G. Elliot Smith (1924) to suggest the following story of the evolution of the cortex. After the demise of the dinosaurs, mammals extended their range to different habitats, including forest trees. Here, they had to become more reliant on vision, and their good sense of smell, which had served them well when they were living on the ground, was no longer as useful, for the informative scents of soil and undergrowth would simply have been blown away. This, he said, freed up resources that could be used for the construction, maintenance, and further evolution of the forebrain. At the time, it was far from clear what these resources might be, but recent discoveries in olfaction suggest an intriguing possibility.

Consider these three facts: (1) In the mouse, more than 1200 genes are involved in coding for all the different olfactory receptor molecules; (2) each receptor cell's axon finds its way to a glomerulus that also receives axons from other receptor cells in the neighborhood containing the same type of receptor molecule, and Singer, Shepherd, and Greer (1995) concluded that this molecule must have a role in axon guidance, as well as in olfactory reception; and (3) in modern primates, many genes for olfactory receptors are not expressed in the olfactory mucosa. These three facts point to the possibility that these genes have retained an axon-guiding role and that this has helped the cortex to resolve its axon-guiding problems. As was emphasized earlier, cortex is characterized by having a vast number of neurons interconnecting its different parts, so it faces this axon-guiding problem on an unparalleled scale.

Whether or not there is any truth in this notion, what one really wants to know about is how their enlarging cortex allowed mammals to enlarge their behavioral repertoire: What could they do with a cortex that they could not do without it? Unfortunately, most of the suggested answers to this question are useless, because they depend upon words or phrases—such as *voluntary actions*, *free will*, or *consciousness*—that we do not understand. In contrast, C. J. Herrick (1926, 1928), a founding father of comparative neuroanatomy in the United States, penned three one-liners that were astonishingly prescient.

First, he said that cortex contained the "filing cabinets of the central executive." He was mainly suggesting that cortex is the store-place of memory, but perhaps he was also anticipating the idea that determining, storing, and making use of the statistics of the environment are crucially important for everything we do, especially for those things that primates, and particularly humans, do better than other animals or computers.

Second, he called cortex the "organ of civilisation." Here again, he anticipates understanding of the fact that civilization is possible because humans are adept at assimilating, disseminating, and using little associative facts, such as "Fanny likes bananas." Substitute other, less trivial nouns and verbs for those in this sentence, then allow that individuals collect and remember associations that are observed to be true, and you have a population with many shared beliefs that can, together with a system of intricately linked mutual benefits, make a civilized society possible.

Neither of these gives much idea of the actual operation cortical neurons perform, but Herrick's third one-liner was "cortex is the organ of correlation." The ability to detect associations underlies the other two, for "Fanny likes bananas" proclaims that you have noticed a correlation between Fanny, the individual, and her manifestation of behavior that shows a liking for bananas; furthermore, most of the facts in those filing cabinets of Herrick's first one-liner are not just the means and current values of a host of variables, but also evidence for associations among them, most of them weightier than that just mentioned.

Correlation, symmetry, and invariance

Herrick did not specify exactly what he meant by calling the cortex the organ of correlation. He might have just meant that it detects the similarity between patterns in the external world and patterns held as templates in the brain. If a receptive field with a particular pattern of spatial sensitivity, say a Gabor function, calculates the point-by-point linear sum of sensitivity times luminance over its field, it becomes a matched detector for that same spatial pattern in an image falling on it. Such a matched detector can give the highest possible signal-to-noise ratio for detecting a precisely known stimulus pattern, and some of the properties of biological feature detectors can be well modeled in this way. Such models have two serious problems, however: They do not show the invariance of response for position, illumination, pose, or size that is typical of biological pattern recognition, and they do not form a good basis for hierarchical models, because the result of doing two (or more) template matching operations in sequence can be expressed as a single template matching operation. There are ways of ameliorating these problems (see Heeger, 1992; Carandini, Heeger, & Movshon, 1997; Simoncelli & Heeger, 1998; Rust, Mante, Simoncelli, & Movshon, 2006), but there is also a more radical solution.

An "organ of correlation" should be able to reveal more about an image than which templates best match its different parts. Autocorrelation is an operation that is related to template matching but differs from it in that it compares parts of an image with other parts of the same image rather than with a fixed template. This has been used to follow the motion of planetary atmospheres (Luz, Berry, & Roos-Serote, 2008), and it allows the symmetries in sensory messages to be actively sought out.

The most appropriate definition of symmetry that we have found is "the property of remaining invariant under

certain changes" (Merriam-Webster online dictionary). For example, a circular object has circular symmetry because it does not change when you rotate it about its center, and a line or edge has translational symmetry because it does not change when you move it in the direction of its orientation. In popular usage, the word *symmetry* most often refers to bilateral, or mirror, symmetry, meaning that if one half of an image is flipped over, it will exactly match the other half onto which it has been flipped, provided that the axis of the flipping has the right position and orientation (i.e., if it is the mirror axis).

The responses of retinal ganglion cells to image contrast are little changed by altering the mean level of illumination. One might call this "contrast symmetry" or "invariance to mean illumination," but note that such changes of mean illumination are passively experienced, not actively sought out. It is not like rotating an object and finding its appearance unchanged, for that involves the active application of the "certain changes."

A way to bring out the importance of symmetry is to point out that it tells you how to find more of the same; it tells you that if symmetry is present and you apply the changes for which invariance holds, you will find more evidence of the same type, to be likely from the same object. Take displacement symmetry, which is closely related to translational symmetry; it is present if you move across an image through a specified distance and direction and find that the region at which you arrive has properties similar to those of the region from which you departed. That can easily be tested for by rotating the eye through the specified distance and direction, when any small patches over the whole image that are unchanged by the eye movement must have the specified symmetry. However it is obvious that at only three saccades per second, the eye cannot quickly determine the displacement symmetries that are present in a whole scene using this method.

From what we know about the brain's computing powers, it is hard to think of any simple alternative to the following brute force general scheme for detecting symmetries: Determine the properties, such as luminance, color, and texture, of each small region, using a limited number of local characteristics; transmit their values to all the neighboring regions for which displacement symmetry might exist; and test the similarities of the received values with those in the receiving small regions. At first, this might seem absurdly extravagant in the use of neurons and likely to require far too many of them, but recall that in the foveal region of V1, there are some 10,000 cortical neurons per input fiber from the LGN. There are a lot of neurons available for the task, but I do not think that the evidence is yet available to make even a crude estimate of whether there are enough.

From the definition, it is clear that searching for symmetry is the same as searching for ordered modifications to the image that leave certain parts of it unchanged. Rotation, translation, and reflection were mentioned above, and there are many other tasks in perception that can be easily framed in terms of symmetry and invariance. For example, when a small region in an image moves, the pixel values of all points in the small region are displaced through the same distance and direction and are also delayed in time, so the displaced and undisplaced small regions of the image have spatiotemporal displacement symmetry; this is the symmetry involved in motion detection. Similarly, the binocular disparity that gives rise to stereoscopic depth perception occurs when a small region of one eye's image is displaced horizontally relative to its position in the other eye's image, and this is the symmetry involved in stereoscopy. It is an interesting exercise to go through the features in images that the Gestalt school pointed to as promoting grouping or segregation, asking, "How are symmetry and invariance involved here?" In most cases, it is not hard to see that there is a connection, though it is sometimes hard to find exactly the right words to express it.

We need to consider with particular care Adelson and Movshon's (1982) plaid paradigm, for this is perhaps the only perceptual phenomenon that has not only been shown to be generated in the cortex (Movshon, Adelson, Gizzi, & Newsome, 1985) but for which there is also a neurophysiological model that works (Rust et al., 2006).

To generate a plaid, two moving gratings that have different spatial frequencies, orientations, contrasts, and velocities are superimposed. Sometimes you see just what you might expect, namely, two different gratings, each moving in its own direction. However, if the two gratings are not too different from each other, you often experience something new, namely, a single plaid rather than two separate gratings, and its direction of movement is obviously different from that of either of the original gratings. In the second of the three papers cited above, the authors reported finding neurons in area MT of macaques that showed the type of directional selectivity required to signal the direction of the plaids. By contrast, in V1 itself, they found no such neurons, only ones that had the type of directional tuning expected for signaling the motion of each component by itself.

In the third paper, Rust and colleagues (2006) show how the plaid type of directional selectivity can be generated by appropriate interactions among inputs from primary visual cortex (V1) that have the normal directional tuning of V1 neurons and cover a wide range of optimal directions and velocities. Both the input neurons from V1 to MT and the neurons in MT itself had linear/nonlinear characteristics and normalization processes based on previous evidence (Heeger, 1992; Carandini et al., 1997). If symmetry and invariance are important in understanding what cortex does, they should also help us to understand this remarkable sequence of observations, experiments, and modeling, and I think they do.

The merging could be described as the ability to perceive the true direction of motion of a textured region in an image, invariantly for differences of texture. This new description does not much help one to understand the mechanism—indeed, it does not even hint at the intersection of constraints idea that Adelson and Movshon (1982) pointed to as the key to how it must be done. Nor does it enable one to predict the interactions Rust and colleagues (2006) discovered or the model by which they explained them. On the other hand, it does neatly summarize the functional role of the merging mechanism, and it points to two other merits of using invariant symmetries as the "words" that a neuron signals, rather than other more or less arbitrary features.

The region where the merging occurs is MT (Movshon et al., 1985), which receives input largely from V1 and transmits output largely to MST. These three areas are all strongly responsive to motion and often strongly selective for its direction, but the V1 neurons give potentially misleading information, because the velocity vector to which they respond best is not invariant for the spatial frequency composition (i.e., texture) of the stimulus. MT corrects for this lack of invariance, and one can see why it is needed by looking at the properties of neurons in MST. These respond maximally to motions in different directions in different parts of their receptive fields, and the patterns of motion that fulfill the requirements all over the receptive field are the patterns of motion, called *optic flow*, that occur when an observer moves through a textured environment. Now with optic flow, the textures of different parts of the scene are likely to be very different, so direct information from V1 to MST would often be misleading. The invariance for texture differences of the MT neurons would not only avoid this, but also allow these neurons to combine information appropriately from a larger number of V1 neurons, thus improving the reliability and signal-to-noise ratio of the messages from MT neurons.

It requires the concepts of symmetry and invariance to understand this neat and economical system, and these concepts might also be needed to explain how the model of Rust and colleagues (2006) is set up in the monkey's area MT. The constraints on patterns of apparent motion in the image that result from geometric optics and a single observing instrument moving through a structured three-dimensional environment are constantly present in the inputs to a monkey's eye. Could the monkey's visual cortex seek out the symmetries and invariances that result from these constraints and use them to adjust the parameters of a genetically prewired skeleton of the model? Einstein plucked his scientific laws from the confusing and sometimes contradictory evidence of contemporary physicists, and we generally accept that his genius lay in his cerebral cortex. Why should not monkey cortex be using a touch of Einstein's genius to generate an economical model of its motion environment?

Such economy has long been thought desirable for sensory systems and perception (Attneave, 1954; Barlow, 1959; Watanabe, 1960); in particular, it can make possible a sparser representation of a scene. The advantages that this brings for improving the reliability and sensitivity of learning and economy of metabolic costs are explained elsewhere (Olshausen & Field, 1997; Gardner-Medwin & Barlow, 2001; Attwell & Laughlin, 2001; Lennie, 2003).

The importance of symmetry in perception and in psychology at all levels needs more attention. At one extreme, rhetoric often uses analogy to strengthen a case, and that depends upon recognizing the symmetry between two arguments, one already accepted and the case for acceptance of the other supposedly being strengthened by the postulated symmetry. At the other extreme, we have the starburst amacrines of the retina, which are believed to be the means by which knowledge of the image at one point is transmitted to another point to provide evidence for spatiotemporal displacement symmetry, the symmetry underlying motion detection (Vaney & Taylor, 2002; Euler, Detwiler, & Dent, 2002).

Symmetry appears to be something that is widely exploited by the brain, and it probably has a role in the mechanisms underlying grandmother cells and face detection. This carries the immediate lesson that it is at least as important to define and puzzle over the invariances of grandmother cells' responses as it is to study their pattern selectivity.

Conclusions

Neurons that respond to "a highly complex, specific, and meaningful stimulus" (Gross, 2002) exist and can now be recorded from and studied reliably. But grandmothers are not among the specific stimuli to which they respond, and their ability to respond invariantly is as important as their pattern selectivity.

An abrupt change in the anatomy of the visual pathway occurs as it enters the cerebral cortex. The total number of neurons in V1 over the whole visual field is about 250 times the total number of LGN neurons, and if one considers the foveal region alone, there appear to be 10,000 cortical neurons per LGN afferent. If cortical computations require this number of neurons, they are likely to be accomplishing more than we currently expect of them. Although the idea of grandmother cells must have been based largely on the recognition that subcortical single neurons are much more pattern selective and reliable than had previously been supposed, it could be misleading to assume that the cortex can do no more than subcortical mechanisms.

Most cortical models have been based on the summed cross-product of the pattern of activity in a group of input messages with the pattern of sensitivity of a receptive field, which thus acts as a weighting function or template. So far,

this has failed to provide an adequate model of the selectivity and invariance of face cells.

It has been suggested that the cortex is "the organ of correlation" (Herrick, 1928) and can detect the invariances of symmetries by autocorrelation. Autocorrelation involves the comparison of two patterns, as with template matching, but both patterns are from the image, instead of one being from the image and the other from an internally stored template.

Autocorrelation would require the transmission of information from one region of an image to other regions on a massive scale, but as was emphasized above, phenomenal numbers of neurons are available to meet this demand.

The ranges of invariance that symmetries define increase the range of stimuli that excite a neuron and thus can improve its signal-to-noise ratio and reliability. Achieving maximum invariance with minimum loss of needed selectivity must be a principle that operates at most, if not all, stages in the successive operations that determine the properties of grandmother cells.

REFERENCES

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300*, 523–525.

Afraz, S.-R., Kiani, R., & Esteki, H. (2006). Microstimulation of infero-temporal cortex influences face categorisation. *Nature, 442*, 692–695.

Attneave, F. (1954). Informational aspects of visual perception. *Psychol. Rev., 61*, 183–193.

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab., 21*, 1133–1145.

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. In *The mechanisation of thought processes* (pp. 535–539). London: Her Majesty's Stationery Office.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1*, 371–394.

Barlow, H. B. (1981). Critical limiting factors in the design of the eye and visual cortex. The Ferrier lecture, 1980. *Proc. R. Soc. Lond. B Biol. Sci., 212*, 1–34.

Barlow, H. B. (1995). The neuron doctrine in perception. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 415–435). Cambridge, MA: MIT Press.

Barlow, H. B., Hill, R. M., & Levick, W. R. (1964). Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *J. Physiol., 173*, 377–407.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci., 17*, 8621–8644.

Chow, K.-L., Blum, J. S., & Blum, R. A. (1950). Cell ratios in the thalamo-cortical visual pathway of *Macaca mulatta*. *J. Comp. Neurol., 92*, 227–239.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci., 4*, 2051–2062.

Elliot Smith, G. (1924). *The evolution of man*. Oxford, UK: Oxford University Press.

Euler, T., Detwiler, P. B., & Denk, W. (2002). Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature, 418*, 845–852.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex, 1*, 1–47.

Gardner-Medwin, A. R., & Barlow, H. B. (2001). The limits of counting accuracy in distributed neural representations. *Neural Comput., 13*, 477–504.

Gross, C. (2002). Genealogy of the "grandmother cell." *Neuroscientist, 8*, 512–518.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in the infero-temporal cortex of the monkey. *Science, 166*, 1303–1306.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in infero-temporal cortex of macaque. *J. Neurophysiol., 35*, 96–111.

Hartline, H. K. (1949). Inhibition of activity of visual receptors by illuminating nearby retinal areas in limulus. *Fed. Proc., 8*, 69.

Hawken, M. J., & Parker, A. J. (1991). Spatial receptive field organisation in monkey V1 and its relationship to the cone mosaic. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 83–93). Cambridge, MA: MIT Press.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*, 2425–2430.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Vis. Neurosci., 9*, 181–197.

Herrick, C. J. (1926). *Brains of rats and men*. Chicago: University of Chicago Press.

Herrick, C. J. (1928). *An introduction to neurology*. Philadelphia: W.B. Saunders.

Hubel, D. H. (1988). *Eye, brain, and vision*. New York: Freeman.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol., 148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *J. Physiol., 160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two non-striate areas (18 & 19) of the cat. *J. Neurophysiol., 28*, 229–289.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extra-striate cortex specialized for face perception. *J. Neurosci., 17*, 4302–4311.

Konorski, J. (1967). *Integrative activity of the brain: An interdisciplinary approach*. Chicago: University of Chicago Press.

Kreiman, G., Koch, C., & Fried, I. (2000). Category specific visual responses of single neurons in the human medial temporal lobe. *Nat. Neurosci., 3*, 946–953.

Lennie, P. (2003). The cost of cortical computation. *Curr. Biol., 13*, 403–407.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proc. Inst. Radio Eng., 47*, 1940–1951.

Levick, W. R. (1967). Receptive fields and trigger features of ganglion cells in the visual streak of the rabbit's retina. *J. Physiol. Lond., 188*, 285–307.

Luz, D., Berry, D. L., & Roos-Serote, M. (2008). An automated method for tracking clouds in planetary atmospheres. *New Astron., 13*(4), 224–232.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern recognition mechanisms* (*Pontificiae Academiae Scientarium Scripta Varia*, Vol. 54, pp. 117–151). Rome: Vatican Press. (Reprinted in *Exp., Brain Res.*, Supplement 11, 117–151, 1986.)

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci., 12*, 1013–1023.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res., 37*, 3311–3325.

Perrett, D. I., Heitanen, J. K., Oram, M., & Benson, P. (1992). Organisation and function of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 335*, 23–30.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Exp. Brain Res., 47*, 329–342.

Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J. Neurophysiol., 74*, 1192–1199.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435*, 1102–1107.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the nervous system. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 303–317). New York: Wiley.

Riesenhuber, M., & Poggio, T. (1994). How the visual cortex recognizes objects: The tale of the standard model. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (Vol. 2, pp. 1640–1664). Cambridge, MA: MIT Press.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nat. Neurosci., 9*, 1356–1357.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Res., 38*, 743–761.

Singer, M. S., Shepherd, G. M., & Greer, C. A. (1995). Olfactory receptors guide axons. *Nature, 377*, 19–20.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? A fMRI study. *Neuron, 35*, 1157–1165.

Tsao, D. (2006). A dedicated system for processing faces. *Science, 314*, 72–73.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nat. Neurosci., 6*, 989–995.

Tsao, D., Freiwald, W. A., Tootell, R. B. H., & Livingston, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science, 311*, 670–674.

Vaney, D. I., & Taylor, W. R. (2002). Directional selectivity in the retina. *Curr. Opin. Neurobiol., 12*, 405–410.

Watanabe, S. (1960). Information-theoretical aspects of inductive and deductive inference. *I.B.M. J. Res. Dev., 4*, 208–231.

Wiersma, C. A. G., Waterman, T. H., & Bush, B. M. H. (1961). The impulse traffic in the optic nerve of decapod crustacea. *Science, 134*, 1453.

Yamani, S., Kaji, S., & Kawano, K. (1988). What facial features activate face neurons in the infero-temporal cortex of the monkey? *Exp. Brain Res., 73*, 209–214.

22 Olfaction: From Percept to Molecule

YAARA YESHURUN, HADAS LAPID, RAFI HADDAD, SHANI GELSTIEN, ANAT ARZI, LEE SELA, AHARON WEISBROD, REHAN KHAN, AND NOAM SOBEL

ABSTRACT Despite major progress in elucidating the anatomical and molecular foundations of olfaction, the rules underlying the link between the olfactory stimulus and the olfactory percept remain unknown. We argue that this lack is a reflection of visual primacy in human perception and thinking, primacy that has prevented the development of a perception-based approach to studying olfactory coding. With this in mind, in this chapter, we first provide a tutorial on the organization of the mammalian olfactory system and then describe our recent efforts to generate a perception-based olfactory metric. The primary olfactory perceptual axis revealed by this effort was odorant pleasantness. We found that pleasantness is a perceptual representation of the physical axis that best explains the variance in molecular odorant structure. That the most important dimension in olfactory perception should be the best correlate of the most discriminating physicochemical measures suggests that, as with other senses, the olfactory system has evolved to exploit a fundamental regularity in the physical world. In this respect, olfactory pleasantness can be likened to visual color and auditory pitch. Finally, we review our use of this olfactory metric in predicting odor perception in humans and odorant-induced neural activity in the olfactory system of nonhuman animals.

Our lab has a pet cat named Diesel. We found Diesel at the age of about four weeks, when he was suffering from severe feline herpes that had invaded both his eyes. Despite significant veterinary efforts, both his eyes had to be removed, and Diesel has since been a blind cat. That said, any naïve visitor to our lab will not notice Diesel's blindness. Diesel runs and jumps around the lab following an internal spatial mental map he has constructed, and he negotiates changes in this landscape, such as a chair that has been moved, with surprising speed, thanks to rapid processing of information from his whiskers, which are always a step in front of him when he is in motion. However, the most astonishing aspect of Diesel's behavioral repertoire is his ability to catch flies in flight. Diesel will identify the sound of a fly seemingly despite any level of background noise, will follow the fly seemingly tracking the fly with his eyes (which are not there), and at the correct moment will leap several feet into the air, catching the fly between his clapped paws (figure 22.1).

Observing this marvelous demonstration of sensory substitution (audition for vision) is a lesson to us in our studies of olfaction. We humans are visual animals, and this has shaped not only how we negotiate the world around us, but also how we think about it. Vision dominates our conscious perception. We intuitively think that information about the outside world that is naturally provided to us through vision is inherently *visual information*. This is not necessarily true, however, as is so powerfully shown in Diesel's fly hunting. In other words, all distal senses have evolved to maximize the amount and types of information they can extract from the environment. It follows from this that if we pay careful attention to our sensory perceptions in each domain, we can learn much about how that domain is physically organized in the world around us, and in our brain. This simple truth was elegantly stated by Helmholtz (1878): "Thus, even if in their qualities our sensations are only signs whose specific nature depends completely upon our make-up or organization, they are not to be discarded as empty appearances. They are still signs of something—something existing or something taking place—and given them we can determine the laws of these objects or these events. And that is something of the greatest importance!"

Neurobiologists have internalized this lesson for vision and audition but not for olfaction. In vision, we learned that what appears to us as perceptually similar in color is in fact similar in the physical dimension of wavelength; and in audition, we learned that what appears to us as perceptually similar in pitch is in fact similar in the physical dimension of frequency. We further learned that the physical dimensions of wavelength and frequency are represented at various stages of the nervous system. But what physical dimension is common to similarly smelling odors? And how is this dimension represented in the brain? These questions remain unanswered. In this chapter, we will first provide a basic tutorial on the organization of the mammalian olfactory system and then describe our initial efforts to generate a perception-based approach for probing the neurobiology of olfaction.

YAARA YESHURUN, HADAS LAPID, RAFI HADDAD, SHANI GELSTIEN, ANAT ARZI, LEE SELA, AHARON WEISBROD, REHAN KHAN, and NOAM SOBEL Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel

FIGURE 22.1 Diesel, a blind cat, aged six months, hunts flies in flight using audition only. Here, Diesel jumps to swat at the slightly moving hand of coauthor N.S.

The olfactory system

What is an odor? Any volatile (molecular weight < 294 daltons) molecular species with surface activity, low polarity, some water solubility, high vapor pressure, and a high lipophilicity will probably be detectable and discriminable by the mammalian olfactory system (Ohloff, 1986). Nobody knows exactly how many discrete odorants this amounts to. Estimates have ranged from thousands to tens of thousands to hundreds of thousands, yet we know of no computed theoretical upper limit to this number. The human olfactory system can detect these molecules with astonishing sensitivity, outperforming analytical instruments (Cain, 1977) and performing on par with monkeys (Laska, Trolp, & Teubner, 1999) (figure 22.2). The transformation from molecule to percept occurs along an hierarchically organized system that in mammals is summarized as follows: Following transduction at olfactory receptor neurons in the olfactory epithelium, odor information is projected ipsilaterally via the olfactory nerve to the olfactory bulb. Following bulbar processing, the signal is projected ipsilaterally via the lateral olfactory tract to primary olfactory cortex within the ventral portions of the temporal lobe (figure 22.3).

In this manner, olfaction differs from the distal senses of vision and audition, where peripheral input projects to contralateral cortex initially via a thalamic relay. In contrast, olfactory information projects to the thalamus from primary olfactory cortex. Additional primary olfactory cortex projections relay odor information to multiple brain regions, including what has been referred to as secondary olfactory cortex in the orbitofrontal region and flavor integration regions in the insula (Small & Prescott, 2005).

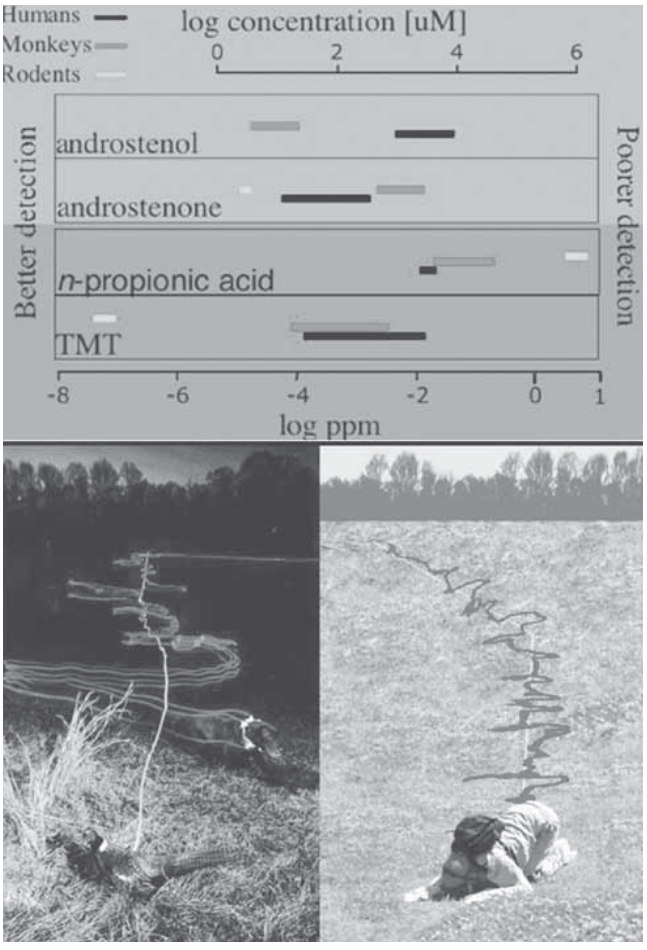


FIGURE 22.2 Olfaction in humans and other animals. *Top panel*: Detection thresholds across species. The data are amassed from studies by Laska and colleagues (1999) and are for detection of the fox odor TMT, n-propionic acid, and the two steroidal compounds androstenol and androstenone. The two shades of gray distinguish the units in which results were reported, dark gray in log concentration of the vapor, and light gray in log concentration of the odorant liquid. The extent of the lines reflects the reported variance across studies. The important point illustrated is that each species excels at detecting particular odorants. For example, humans outperform rats and monkeys at detecting n-propionic acid. That said, one must keep in mind the limitation of comparing across studies that used different methods of delivery and statistical criteria. *Bottom panel*: Human subject's path following a scent trail, as compared to a dog's path. *Left*: Path of a dog following the scent trail of a pheasant dragged through a field (scent trail in yellow, dog's path in red) (Gibbons, 1986). *Right*: Path of a human following a scent trail of chocolate essential oil through a field (scent trail in yellow, human's path in red). (The background trees were pasted in for esthetics and are part of the data.) (From Porter & Sobel, 2005.) (See color plate 22.)

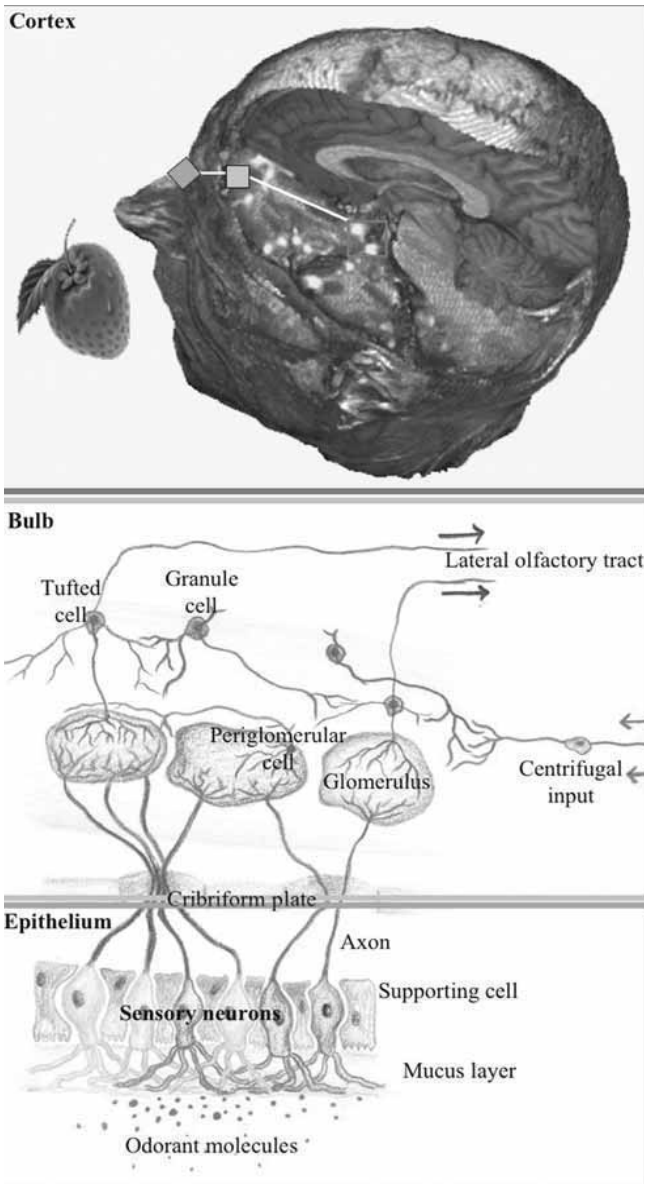


FIGURE 22.3 Structure of the human olfactory system. The human olfactory system can be segregated into three primary compartments: (bottom) epithelium, (middle) bulb, and (top) cortex. Olfactory epithelium: Each olfactory sensory neuron expresses on olfactory receptor gene. Like receptor project to one or a small number of glomeruli. Organization of the olfactory bulb: Glomeruli receive input from olfactory sensory neurons and cortical olfactory regions. Mitral and tufted cell dendrites contact receptor axons within glomeruli. The axons of the mitral and tufted cells project widely to higher brain structures. Lateral processing in the olfactory bulb occurs across two types of interneurons: periglomerular cells and granule cells. A sagittal view of the human head. The olfactory epithelium is in green, bulb in blue, and primary cortex in pink. (Drawing courtesy of Christina Zelano.) (See color plate 23.)

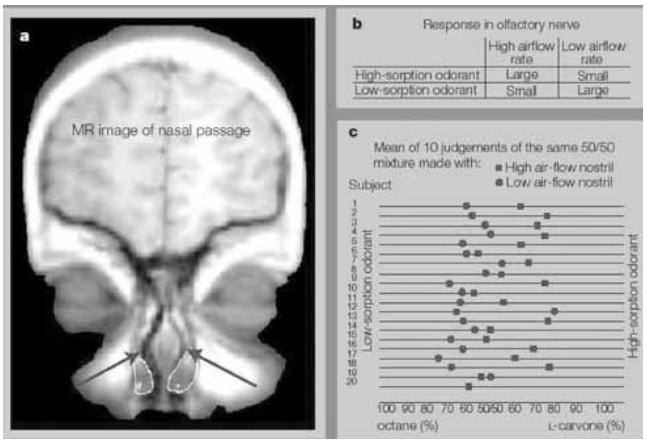


FIGURE 22.4 Different nostrils convey different olfactory information to the brain. (*A*), Magnetic resonance image of the nasal passages. The swollen (*) and relaxed (#) turbinates, outlined in white, result in an occluded right nostril (red arrow) and a clearer left nostril (green arrow). (*B*), The size of the response in the olfactory nerve (large or small) as a function of the interaction between airflow rate and odorant sorption (Mozell & Jagodowicz, 1973). (*C*), On each of 10 trials, subjects were asked to smell an identical mixture of 50% octane and 50% L-carvone using either the left or right nostril. They were then given each individual odorant component to smell separately and judged the composition of the mixture by marking the line. Using the high-flow-rate nostril (green), the average judgment was that the mixture consisted of 55% L-carvone and 45% octane. Using the low-flow-rate nostril (red), the judgment was that it consisted of 61% octane and 39% L-carvone ($t(19)43.74$, $p = 0.001$). (From Sobel et al., 1999.) (See color plate 24.)

The above is a unilateral description of what is a bilateral system. As in other distal senses, mammals have two olfactory systems, a left and a right nostril, each connected to their respective epithelia, bulbs, and cortices. Also as in other distal senses, there is an asymmetry of input across sides, whereby the left and right nostril are tuned to slightly different aspects of the olfactory content (Sobel, Khan, Saltman, Sullivan, & Gabrieli, 1999). Whereas this asymmetry may contribute to spatial localization of odors (Porter, Anand, Johnson, Khan, & Sobel, 2005; Rajan, Clement, & Bhalla, 2006; Porter et al., 2007), any contribution of this sensory offset to olfactory discrimination remains unknown (figure 22.4).

Airborne chemicals are concurrently transduced by at least two additional mammalian neural subsystems: the vomeronasal system (Halpern, 1987; Meredith, 1991; Keverne, 1999; Dulac, 2000), the trigeminal system (Doty, 1995; Hummel, 2000; Hummel & Livermore, 2002), and possibly the Grüneberg organ, a septal subsystem of the main olfactory system (Roppolo, Ribaud, Jungo, Luscher, & Rodriguez, 2006; Storan & Key, 2006). However, to maintain a manageable scope in this chapter, we will focus on the main olfactory system only. We will first briefly summarize

the structure and events at each of the olfactory anatomical processing stages; then we will review the current understanding of how odor is encoded within this neuronal infrastructure.

Olfactory epithelium

For a molecule to be transduced into a neural signal, it must reach olfactory receptors within the olfactory epithelia. The human epithelia are located bilaterally about 7 cm up the nasal passage, lining the cribriform plate and extending to the nasal turbinates (Clerico, To, & Lanza, 2003). These turbinates are convoluted to increase the surface area of the epithelium to about 1–10 cm^2 in humans (Moran, Rowley, Jafek, & Lovel, 1982) and 170 cm^2 in some dogs (Moulton, 1977). An odorant molecule may reach the epithelium in three different ways. The first is by diffusion from an area of higher concentration (the environment) to an area of lower concentration (the nares). The second is by a process termed *retronasal olfaction* (Hornung & Enns, 1986), whereby an odorant enters the mouth in food or drink and propagates back up the throat into the nose. The third and most common route by which a molecule reaches the olfactory epithelium is transportation by either ongoing nasal inhalation or a sniff—a vigorous contraction of the diaphragm leading to rapid nasal airflow (often exceeding 100 liters per minute in humans) (figure 22.5). Sniffs are a rapidly modulated mechanism of sensory acquisition that profoundly influence patterns of neural activity throughout the olfactory system and have a significant impact on the olfactory percept (Kepecs, Uchida, & Mainen, 2006, 2007; Mainland & Sobel, 2006; Schoenfeld & Cleland, 2006).

The epithelium consists of many cell types (Carr, Farbman, Colletti, & Morgan, 1991; Huard, Youngentob, Goldstein, Lauskin, & Schwob, 1998), but most fall into four primary categories: olfactory receptor cells, sustentacular or supporting cells, basal cells, and duct cells of Bowman's glands. The latter are the secretory source of a mucus layer that lines the olfactory epithelium. This mucus layer plays a role in immune function (Getchell & Getchell, 1991) and various enzymatic processes (Lewis & Dahl, 1995) but also directly affects olfaction by selectively modulating the passage of odorants to the receptors and possibly by further modulating the later removal of odorants from the receptors (deactivation) (Lewis & Dahl, 1995). In addition to the passive mucosal effect on olfaction, the mucosa contains an odorant-binding protein that assists in actively transferring hydrophobic odorants across the largely hydrophilic mucosa (Pevsner, Trifiletti, Strittmatter, Sklar, & Snyder, 1985; Pelosi, 2001).

The olfactory receptor cells are bipolar neurons that are unique in at least two ways: (1) They constantly regenerate from the basal cell layer (Graziadei, Levine, & Monti Graziadei, 1979; Graziadei & Monti Graziadei, 1983), with a typical life span in mammals ranging from a month to a year (Hinds, Hinds, & Mcnelly, 1984; Mackay-Sim & Kittel, 1991), and (2) they are in direct contact with the external environment. This direct link of the brain to the outside world has been postulated as a path of entry for pathogens from the environment directly to the brain (Roberts, 1986). Humans have about six million receptor cells in each epithelium (Moran et al., 1982). Each of these olfactory receptor neurons sends one dendrite into the mucus layer, terminating in an olfactory knob that contains between 3 and 50 nonmotile olfactory cilia, each about 5 μm long (Morrison & Costanzo, 1990, 1992). It is these cilia that contain the site of olfactory transduction. Various lines of evidence suggested that this transduction was similar to transduction of light in retinal rods and cones. Doron Lancet and colleagues found that odorants induce activation of adenylate cyclase (Pace, Hanski, Salomon, & Lancet, 1985), and Nakumara and Gold (1987) found a cyclic nucleotide–gated conductance in olfactory receptor cilia. Randy Reed and colleagues later found this adenylate cyclase to be olfactory-specific (Bakalyar & Reed, 1990) and activated by an olfactory-specific G protein (Jones & Reed, 1989). In 1991, Linda Buck and Richard Axel culminated this line of research by identifying a large multigene family that encodes the olfactory receptors. The olfactory receptors indeed belonged to the family of G-protein-coupled receptors of the seven-helix type, that, surprisingly, contained around 1000

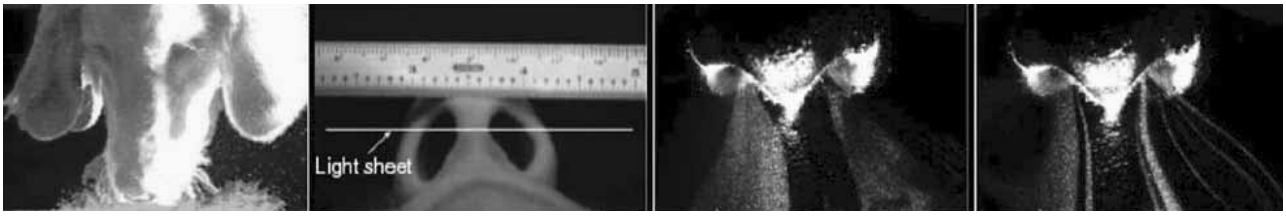


FIGURE 22.5 Airflow visualization using Schlieren imaging in dogs (Settles, 2001) and humans (Porter & Sobel, 2005). *Left panel*: Before sniffing inward, dogs may also sniff outward in a lateral trajectory distributing particles so that they can be inhaled and smelled. As this image clearly depicts, sensation is an active process, and olfaction is a good case in point. *Remaining panels*: Imaging of the human nose clearly reveals an asymmetry in airflow into each nostril.

different olfactory receptor genes, the largest known gene family in the mammalian genome (Buck & Axel, 1991). This work has generated the following current view of olfactory transduction: When an odorant molecule binds to an olfactory receptor, it triggers the activation of the specific olfactory G-protein G_{olf} that releases a subunit $G_{\alpha olf}$ that stimulates adenylyl cyclase III. Adenylyl cyclase III increases intracellular cAMP, which opens a cyclic nucleotide–gated cation channel, depolarizing the cell and ultimately resulting in an action potential in the sensory neuron (figure 22.6). John Ngai and colleagues demonstrated that blocking this cascade causes anosmia—a total loss of olfaction (Brunet, Gold, & Ngai, 1996).

A second transduction cascade has also been suggested whereby activation of the G-protein and phospholipase C lead to the production of IP_3, which directly opens calcium channels that depolarize the cell (Boekhoff, Tareilus, Strotmann, & Breer, 1990). Following transduction, odorant-induced action potentials propagate down the axons of olfactory receptor neurons that join to form the olfactory nerve, pass through the cribriform plate, and synapse at the olfactory bulb.

Olfactory bulb

The mammalian olfactory bulb consists of six cellular layers, which are, from superficial to deep, the (1) olfactory nerve layer, (2) glomerular layer, (3) external plexiform layer, (4) mitral cell layer, (5) internal plexiform layer, and (6) granule cell layer (Kratskin & Belluzzi, 2003). These layers are arranged in a concentric manner reminiscent of an onion (Shepherd, 1972; Greer, Stewart, Kauer, & Shepherd, 1981). The bulb contains functional elements primarily consisting of an input neuron, an interneuron, and an output neuron. The inputs are from two sources: peripheral input from the olfactory receptors and centrifugal inputs from cortical olfactory regions (see figure 22.3). These centrifugal inputs are very extensive, are from various olfactory regions, and,

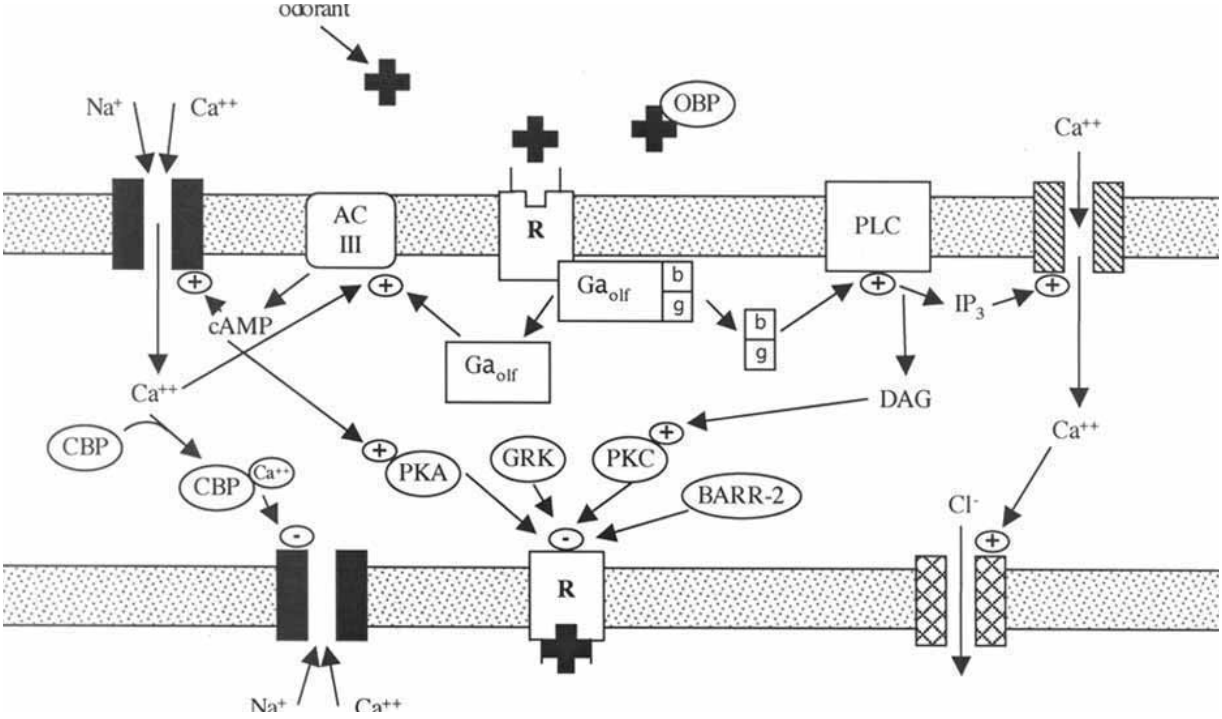


FIGURE 22.6 Odorant molecules bind to olfactory receptors (R) embedded within the olfactory epithelium. The binding causes the associated G-protein complex to release its two subunits (α and β). The α subunit stimulates the integral membrane protein adenylyl cyclase (AC III), which in turn increases the concentration of cAMP. Cyclic-gated nucleotide (CNG) channels open with cAMP, leading to membrane depolarization; if there is sufficient depolarization an action potential is generated in the sensory axon. The $\beta\gamma$ complex released from the G protein stimulates phospholipase C (PLC) leading to higher intracellular inositol triphosphate (IP_3) and diacylglycerol (DAG). IP_3 opens Ca^{++} channels, allowing Ca^{+} to enter the neuron. The Ca^{+} ions have multiple effector pathways. Ca^{+} stimulates Cl channels, allowing Cl ions to exit the cell (intracellular Cl concentration is greater than extracellular concentration). This ion exchange further depolarizes the membrane. Calcium also inhibits the transduction by combining with Ca^{+}-binding protein (CBP) that closes the CNG channels, ending signal transduction. Signal termination is also mediated by a variety of protein kinases (PKA, GRK, PKC) that phosphorylate the olfactory receptor and by β-arrestin-2 (BARR-2) interacting with the olfactory receptor. Odorant-binding proteins (OBP) in the nasal mucosa may increase odorant solubility and/or receptor-binding affinity, aiding transduction, or may assist in odorant clearance, aiding signal termination. (Image after Buck, 1996.)

although not yet fully understood, are thought to play a significant role in olfaction (Gray & Skinner, 1988; Singer, Kim, Zochocoski, 2007). The peripheral input consists of axons of olfactory receptor neurons. These project unbranched from the epithelium to the olfactory nerve layer of the bulb, where they terminate in spherical neuropil structures 50–200 μm in diameter, called *glomeruli*. These form the glomerular layer, which is one or two glomeruli thick. In humans, each olfactory bulb contains approximately 8000 glomeruli (Meisami, Mikhail, Baim, & Bhatnagar, 1998), and in a striking case of neural convergence, all the axons of same olfactory receptors converge to one glomerulus. (Convergence is even greater in macrosmatic mammals.) Within the glomeruli, the receptor axons contact dendrites of either mitral or tufted output neurons and periglomerular interneurons. The bulb is the site of extensive olfactory processing modulated by interneurons consisting of short-axon cells, inhibitory periglomerular cells, and inhibitory axonless granule cells. Granule cells make inhibitory dendrodendritic reciprocal synaptic connections with mitral and tufted cells. Periglomerular cells project a primary dendrite into glomeruli, where they synapse with sensory axons, and additional dendrites make inhibitory dendrodendritic synapses with mitral and tufted cells. The mitral and tufted cell axons join to form the lateral olfactory tract, which is the output from the bulb to primary olfactory cortex in the ventral portions of the temporal lobe (see figure 22.3). Similar to the olfactory epithelium, the olfactory bulb is characterized by relatively high levels of neurogenesis (Pagano et al., 2000; Bedard & Parent, 2004), mainly in the glomerular layer (Ninkovic, Mori, & Gotz, 2007).

Olfactory cortex

By current definition, primary olfactory cortex consists of all brain regions that receive direct input from the mitral and tufted cell axons of the olfactory bulb (Allison, 1954; Price, 1973, 1987, 1990; de Olmos, Hardy, & Heimer, 1978; Carmichael, Clugnet, & Price, 1994; Shipley, 1995; Haberly, 2001). These make up most of the paleocortex, including (by order along the olfactory tract) the anterior olfactory cortex (also referred to as the *anterior olfactory nucleus* [Brunjes, Illig, & Meyer, 2005]), the ventral tenia tecta, the anterior hippocampal continuation and indiusium griseum, the olfactory tubercle, the piriform cortex, the anterior cortical nucleus of the amygdala, the periamygdaloid cortex, and the rostral entorhinal cortex (Carmichael et al., 1994). In rodents, the anterior olfactory cortex may play a role in interhemispheric communication between olfactory bulbs (Cleland & Linster, 2003), but in humans, both the anterior olfactory cortex and the olfactory tubercle are poorly defined. Piriform cortex, the largest component of primary olfactory cortex in mammals, lies along the olfactory tract at the junction of temporal and frontal lobes and continues onto the dorsomedial aspect of the temporal lobe. Piriform cortex is three-layered allocortex, with a superficial plexiform layer I and pyramidal cells densely packed in layer II and less so in layer III. Projections from the bulb synapse onto dendrites of pyramidal cells within layer I of piriform cortex. Caudally, piriform cortex fuses into the anterior cortical nucleus of the amygdala. Olfactory bulb projections terminate densely on the periamygdaloid cortex, which inhabits the medial surface of the amygdala, and less so on the rostral portions of the entorhinal cortex. The entorhinal cortex is the only portion of primary olfactory cortex that is six-layered and is thus considered transitional between olfactory allocortex and neocortex. The entire cortical complex that forms primary olfactory cortex is extensively interconnected by association fibers projecting across regions to synapse at layer I and within regions traversing the three cortical layers (Shipley, 1995). Furthermore, as was previously noted, the primary olfactory cortical regions project extensive centrifugal input back to the olfactory bulb (Carmichael et al., 1994). Beyond these primary regions, olfactory information is projected throughout the brain, most prominently to orbitofrontal gyri and the insular cortex.

As can be appreciated by both the sheer area and the diversity of cortical real estate that is considered primary olfactory cortex, this definition is far from functional. The term *primary* typically connotes basic functional roles such as early feature extraction, yet as can be expected, a region comprising piriform cortex, the amygdala, and the entorhinal cortex is involved in far more complex sensory processing than mere early feature extraction. It is for this reason that this definition is nearing abandonment (Haberly, 2001; Sobel, Johnson, Mainland, & Yousem, 2003). Indeed, in a thorough review of central olfactory structures, Cleland and Linster (2003) simply shifted the definition by referring to the classical primary olfactory structures as secondary olfactory structures, noting that as Haberly (2001) suggested, the definition of mammalian primary olfactory cortex may better fit the olfactory bulb than piriform cortex.

Odor encoding: From molecule to percept

The anatomy described above consists of three primary compartments: epithelium, bulb, and cortex (see figure 22.3). Some aspect of odor encoding occurs at each one of these processing stages, and it is the combination of activity across these regions that gives rise to the complex percept of odor.

As noted earlier, the number of discrete odorants that mammals can discriminate is unknown. That said, there is uniform agreement that it is more than 1000 (the estimated upper limit number of mammalian olfactory receptor types),

and therefore odorants are probably not encoded in a simple one-receptor-to-one-odorant scheme. There are several lines of evidence suggesting that each olfactory receptor neuron expresses only one (Nef et al., 1992; Strotmann, Wanner, Krieger, Raming, & Breer, 1992; Ressler, Sullivan, & Buck, 1993; Vassar, Ngai, & Axel, 1993; Chess, Simon, Cedar, & Axel, 1994) or two (Goldman, Van der Goes van Naters, Lessing, Warr, & Carlson, 2005) types of olfactory receptors. Furthermore, each individual receptor neuron will respond to multiple odorants, and a given odorant will activate several different olfactory receptor types (Sicard & Holley, 1984; Firestein, Picco, & Menini, 1993; Raming et al., 1993; Sato, Hirono, Tonoike, & Takebayashi, 1994; Krautwurst, Yan, & Reedy 1998; Zhao et al., 1998; Malnic, Hirono, Sato, & Buck, 1999). Thus odor encoding may be the result of a combinatorial scheme in which different receptors respond to different molecular aspects of an odorant and a given odorant is then represented by the subset of receptors that it activates (Axel, 1995; Buck, 1996; Mombaerts, 1999; Firestein, 2001). Such a scheme would enable the encoding of a very large number of different odorants within what is potentially a 1000-dimensional space.

Under the assumption that this is indeed the manner in which the peripheral olfactory system initially transduces odorants, one may ask how this high-dimensional information is organized by the brain to enable odor discrimination. Some unique principles may apply to odor encoding. First, odor coding appears to be synthetic (Wilson & Stevenson, 2006; Stevenson & Wilson, 2007). Most environmentally relevant odorants are complex mixtures of molecules yet are processed as a whole. Indeed, the identification of specific odorant molecules or molecular features is far less behaviorally relevant than is the ability to distinguish the complex odor of a ripened fruit or of a predator from background odors. Anterior piriform cortex may play a role in this process, as it acts as a filter that is driven most strongly by changing stimuli, providing a potential mechanism for olfactory figure-ground segmentation and selective reading of olfactory bulb output (Kadohisa & Wilson, 2006a). Consistent with this, mammals are quite bad at determining the composition of mixtures when the number of components reaches four or higher (Livermore & Laing, 1996; Linster & Smith, 1997). Second, in comparison to vision and audition, odor transduction is relatively slower, because molecules must sorb across the mucosa, a process that takes about 150 ms (Firestein & Werblin, 1989). That the neural substrates of olfaction do not have to fire continuously to keep up with peripheral transduction (in contrast to vision and audition) may enable the system to use the temporal encoding domain to encode complex odorant features. The brain can then combine temporal encoding with spatial encoding in the construction of odor space.

Spatial coding

Several lines of evidence point to spatial encoding in odor discrimination. This spatial coding may take form at the level of the epithelium, at the level of the bulb, and perhaps at the level of cortex. The olfactory epithelium is spatially segregated into four zones roughly along the longitudinal axis of the cribriform plate (Buck, 1996). Although some olfactory receptors are randomly represented across these four zones (Strotmann et al., 1992, 2000; Sullivan, Adamson, Ressler, Kozak, & Buck, 1996), most neurons that express a given olfactory receptor type are restricted to one of the four zones, where they are interspersed with other receptors in a seemingly unordered manner (Ressler et al., 1993; Vassar et al., 1993; Strotmann et al., 1994). Thus a rudimentary spatial map of odor identity is generated across these four zones of the epithelium (Mustaparta, 1971; Moulton, 1976; Thommesen & Doving, 1977; Mackay-Sim, Shaman, & Moulton, 1982; Edwards, Mather, & Dodd, 1988; Kent & Mozell, 1992; Scott, Shannon, Charpentier, Davis, & Kaplan, 1997; Scott & Brierley, 1999). Whereas the epithelium may offer crude spatial ordering, the nature of convergence from epithelium to bulb (Miyamichi, Serizawa, Kimura, & Sakano, 2005) enables a much more detailed spatial map of odor at the bulbar level (figure 22.7). Receptor neurons expressing the same type of receptor typically converge onto two individual glomeruli in the bulb, one on the lateral surface and one on the medial surface, in a manner that is symmetrical between bulbs and similar across individuals within a species (Ressler, Sullivan, & Buck, 1994; Vassar et al., 1994; Mombaerts et al., 1996; Tsuboi et al., 1999). Thus a strict relationship is maintained whereby a single receptor type is represented by a two glomeruli. Considering that different receptors may be sensitive to different aspects of odor molecules (termed *odotopes* or *pharmacophores*), a given odorant may then be represented by a spatial pattern of glomerular activity (Lancet, Greer, Kauer, & Shapherd, 1982; Shepherd, 1985; Johnson & Leon, 2007). Furthermore, lateral inhibition between glomeruli may serve to further sharpen this glomerular spatial map of odor (Mori & Shepherd, 1994; Yokoi, Mori, & Nakanishi, 1995). Several lines of evidence support this notion of odor maps on the olfactory bulb, including 2-deoxyglucose uptake (Stewart, Kauer, & Shepherd, 1979; Jourdan, Duveau, Astic, & Holley, 1980; Johnson, Woo, & Leon, 1998; Johnson, Woo, Hingco, Pham, & Leon, 1999; Johnson & Leon, 2000a, 2000b), electrophysiology (Shepherd, 1985; Mori & Yoshihara, 1995), gene expression (Guthrie, Anderson, Leon, & Gall, 1993; Guthrie & Gall, 1995; Johnson, Woo, Duong, Nguyen, & Leon, 1995; Sallaz & Jourdan, 1996; Inaki, Takahashi, Nagayama, & Mori, 2002), and optical imaging (Kauer, 1988; Rubin & Katz, 1999; Uchida, Takahashi, Tanifuji, & Mori 2000; Meister & Bonhoeffer, 2001).

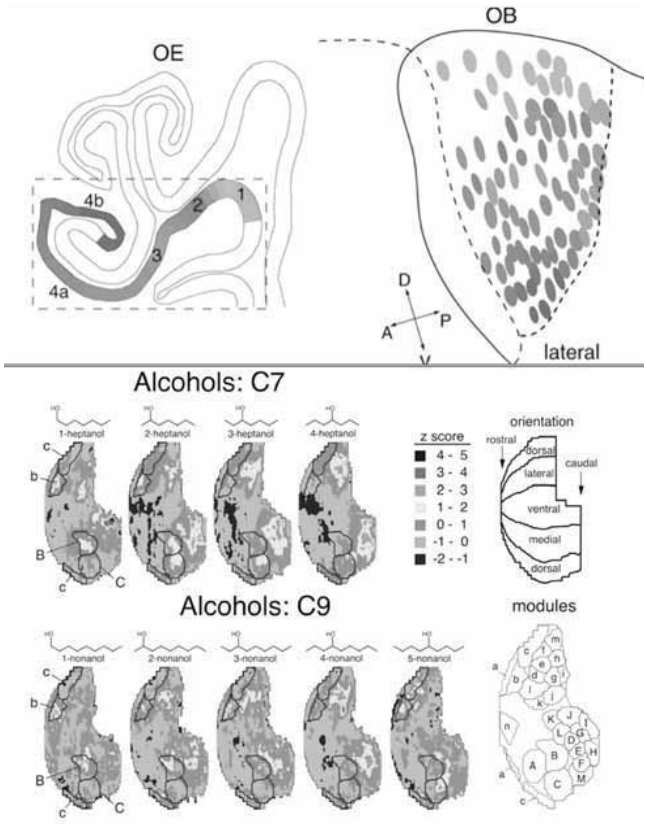


FIGURE 22.7 Spatial mapping from epithelium to bulb. *Top panel*: Color-coded zonal distribution of receptor types in the olfactory epithelijm and their projection pattern to glomeruli in the olfactory bulb (Miyamichi et al., 2005). *Bottom panel*: Patterns of 2DG activation on the surface of the rat olfactory bulb as a reflection of odorant identity. For detailed maps, see http://leonserver.bio.uci.edu/. (See color plate 25.)

If a rudimentary spatial map of odorant identity is formed at the level of the epithelium and then sharpened at the level of bulb, one may ask whether this map is maintained at the level of cortex. Early anatomical studies suggested that the topography of projections from bulb to cortex was disordered and that cortical representation did not maintain the spatial ordering of bulbar representation (Price & Sprich, 1975). However, several functional studies point to spatial axes of representation in primary olfactory cortex. Optical recording of activity in the rat cortex suggest that the piriform cortex is divided in to several functionally heterogeneous regions (Litaudon, Datiche, & Cattarelli, 1997). Electrophysiological recordings suggest piriform cortex has at least four spatial receptive fields specialized for odors delivered to the ipsilateral nostril, contralateral nostril, either nostril, or both nostrils (Wilson, 1997, 2001). Furthermore, a converging finding across rats (Kadohisa & Wilson, 2006b) and humans (Gottfried, Winston, & Dolan, 2006) suggests a dissociation whereby odorant structure is represented primarily in anterior piriform cortex, and odor quality is represented primarily in posterior piriform cortex. Although these findings suggest some relationship between the spatial organization of piriform cortex and odor input, rules underlying this organization remain unknown. Finally, work using single-cell recordings in monkeys found a high degree of odorant-specific responses at an even later stage of olfactory processing, namely, prefrontal and orbitofrontal cortex (Tanabe, Iino, Oshima, & Takagi, 1974; Rolls, Critchley, & Treves, 1996). Taken together, these findings suggest that odor identity may be spatially encoded at the level of cortex, but additional studies are needed to fully address this question.

Under the assumption of spatial mapping of odor at the olfactory epithelium, bulb, and cortex, a key question is what is the metric of smell that governs olfactory perception and hence this map. To date, scientists have been probing for this metric by systematically varying known aspects of molecular odorant structure while imaging the glomerular layer of the bulb. On the basis of work on the I7 receptor in mice and rats (Ivic, Zhang, Zhang, Yoon, & Firestein, 2002), the most commonly probed structural axis is that of carbon chain length. For example, Johnson and colleagues (1999), using 2-deoxyglucose uptake, and Rubin and Katz (1999), Uchida and colleagues (2000), and Meister and Bonhoeffer (2001), using optical imaging, all stimulated with an homologous series of n-aliphatic aldehydes varying in carbon chain length as a possible odor metric. They found that each individual odorant could be represented by a unique spatial distribution of glomerular activity and that the difference between these unique distributions or maps was greater for odorants that differed by increased length of carbon chain. In other words, odorants that differed by one carbon in chain length produced similar glomerular maps, and odorants that differed by several carbons produced increasingly different glomerular maps. Johnson and Leon (2000a, 2000b) defined clusters of glomeruli with similar response specificities as *glomerular modules* and found that in addition to carbon chain length, modular glomerular maps also encode for hydrocarbon branch structure and oxygen-containing functional group. These glomerular module maps were sufficiently consistent to allow Johnson and colleagues (2002) to predict both odorant structure based on patterns of glumerular activity and patterns of glomerular activity in response to novel odorants. The optical imaging studies also found that higher odorant intensities activated increased overall numbers of glomeruli, thus producing a somewhat smeared, yet retained, odorant identity map. This increased recruitment of glomeruli with increased odorant concentration was not always seen with 2-deoxyglucose uptake.

Although the emerging picture of spatial odor encoding at all levels of the olfactory system is quite convincing, this view does have some significant caveats. First, as noted by Laurent (2002), even if the current view of encoding at the

bulb is correct, this is not spatial encoding per se. In audition and vision, spatial ordering in the neural system represents spatial ordering in the physical world. For example, neighboring locations on the basilar membrane encode for neighboring frequencies. This was not the case in the aforementioned studies of activity in the olfactory bulb (Luo & Flanagan, 2007). Second, extensive lesions both at the level of olfactory bulb (Hudson & Distel, 1987; Lu & Slotnick, 1998; Slotnick & Bodyak, 2002) and at the level of the olfactory cortex (Slotnick & Berman, 1980; Slotnick & Risser, 1990; Slotnick & Schoonover, 1993) do little to distort olfactory perception. If this perception were strongly linked to spatial representation within this neural architecture, one would expect these lesions to have far greater impact. Third, odor discrimination can be achieved at a rate that precedes full development of bulbar spatial maps (Wesson, Carey, Verhagen, & Wachowiak, 2008). Finally, patterns of activity throughout the olfactory system in general, and specifically in the olfactory bulb, are constantly modified through experience (Kauer, 1974; Meredith & Moulton, 1978; Di Prisco & Freeman, 1985; Harrison & Scott, 1986; Wellis, Scott, & Harrison, 1989; Eeckman & Freeman, 1990). Pure spatial mapping is an encoding scheme that would probably not lend itself to rearrangement within short time frames. By contrast, temporal encoding of information, namely, encoding of odor within the temporal order of neural activity, may be far more plastic on short time scales and conceivably less susceptible to anatomically restricted lesions.

Temporal coding

Olfactory behavior and therefore neural activity within the olfactory system are marked by rhythmic events. Since the pioneering work of Adrian (1942) and Freeman (1960), it has been known that two particular frequency domains dominate activity throughout the olfactory system. The first is the slow θ rhythm (typically 3–12 Hz) related to sniffing, and the second is the γ rhythm (typically 30–100 Hz), an odor-related oscillation that rides on the respiratory wave. These oscillations occur in both the olfactory bulb and in primary olfactory cortex in a correlated manner (Eeckman & Freeman, 1990) and are present in rodents (Adrian, 1942; Ueki & Domino, 1961; Bressler & Freeman, 1980; Ketchum & Haberly, 1993; Protopapas & Bower, 2001) and perhaps in humans (Hughes, Hendrix, Wetzel, & Johnston, 1969; Sobel et al., 1998). Although it is clear that these oscillations are key to the process of olfaction, the exact functional significance of these patterns remains unclear. Several researchers have suggested that the γ frequencies, coupled with the exceedingly large number of backprojecting pyramidal axons from anterior piriform cortex to the bulb, categorize olfactory stimuli in an increasingly specific fashion over successive sniff cycles (Bressler, 1990; Freeman & Barrie, 1994; Bhalla & Bower, 1997). Consistent with this, single-cell recording revealed that ensemble activity of mitral cells in the olfactory bulb contains information at different timescales that could be separately exploited by downstream brain centers to make odor discriminations (Bathellier, Buhl, Accolla, & Carleton, 2008). Although temporal ordering of neural activity is a well-described encoding strategy in the olfactory system of insects (Laurent, 2002), the suggestion of temporal coding in the mammalian system was met with hesitation by a field that has been dominated by the notion of spatial encoding. That said, studies by Hartwig Spors (Spors & Grinvald, 2002; Spors, Wachowiak, Cohen, & Friedrich, 2006) begins to bridge the gap between these two schools (Friedrich, 2002). Spors and Grinvald (2002) combined optical imaging with voltage-sensitive dyes to obtain high spatial (10–20 μm) and temporal (50–200 Hz) resolution measurements from the olfactory bulbs of rodents. Using these methods the authors found odorant-specific glomerular modules of activity similar to those previously described with optical imaging. However, the added temporal resolution revealed a highly dynamic spatial representation across the glomeruli that was constantly modified both within a sniff and across consecutive sniffs (figure 22.8). In other words, odor was represented by a combined spatiotemporal pattern of activity. This notion is consistent with

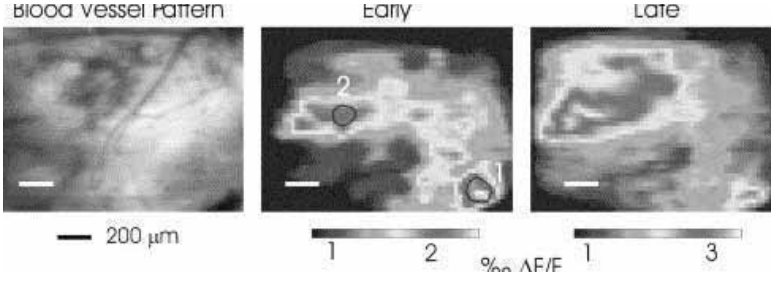


Figure 22.8 Temporal development of odor-induced activity. Data from Spors and Grinvald (2002) showing the temporal development of the bulbar response to the odorant ethylbutyrate. The early response is data obtained 150–300 ms following stimulus onset, and the late response is data obtained 300–500 ms following stimulation. The spatial pattern of response is clearly modified over time. (See color plate 26.)

measurements obtained in cortex as well (Rennaker, Chen, Ruyle, Sloan, & Wilson, 2007). The brain may then read this spatiotemporal pattern as a sequence of discreet spatiotemporal events like frames in a movie (Hopfield & Brody, 2001; Friedrich, 2002). By contrast, these successive representations may represent stages of an ongoing computation in the bulb, in which case "time" is a computational variable in the construction of the odor representation at the bulbar level (Bhalla & Bower, 1997; Friedrich & Laurent, 2001). In both cases, the temporal information on odor may be carried by the intrinsic oscillations whereby it would alter the phase of activity in a specific manner (Hopfield, 1995).

In addition to directly participating in encoding of odorant content, the rapid oscillations may reflect the organization of network activity (Wilson & Bower, 1992; Protopapas & Bower, 2001). Specifically, current source density analysis suggests that each γ oscillation decomposes each inspiratory cycle into temporal bins of about 20-ms duration (Ketchum & Haberly, 1993). Haberly (1998) suggested that the olfactory system uses these temporal bins to pair afferent input from the olfactory bulb with intrinsically generated associational activity and inhibition in piriform cortex in order to subserve this region's primary function as odor association cortex (Haberly, 1985, 2001).

Odor encoding: From percept to molecule

Although it is quite clear that odor is reflected in unique spatial and temporal patterns of neural activity in the olfactory system, odor encoding is far from understood. This is primarily evident in what remains the key unanswered question in olfaction: There is not a perfumer or scientist who can predict the odor of a novel physicochemical structure or predict the physicochemical structure of a novel smell. In our view, olfactory coding has not been solved because the stimulus space is poorly understood. To probe neural coding in vision and audition, neurobiologists varied stimuli along known primary perceptual axes such as color and pitch that were clearly linked to physical dimensions of wavelength and frequency, but what is a primary axis of olfactory perception? And to what physical dimension might it be linked? It is these two questions that we recently set out to ask, and we went about this by shifting from the typical approach of "from molecule to percept" to one "from percept to molecule."

Constructing a perception-based odor space

To build a perception-based odor space, Rehan Khan and colleagues in our lab (Khan et al., 2007) used a previously published data set known as Dravnieks' *Atlas of Odor Character Profiles,* in which approximately 150 experts (perfumers and olfactory scientists) ranked (from 0 to 5, reflecting "absent" to "extremely" representative) 160 odorants (144 monomolecular species and 16 mixtures) against each of the 146 verbal descriptors (Dravnieks, 1985). Dravnieks had demonstrated that this type of data was stable across a large pool of subjects and across a large geographic span (Dravnieks, 1982). Using the data in the atlas, we applied principal components analysis (PCA), a well-established method for dimension reduction that generates an orthogonal basis set for the profile space in which each successive dimension has the maximal possible variance. In simple terms, PCA generates a new set of features, each of which is a weighted linear combination of the input feature space, such that the new features are (1) orthogonal and (2) ordered so that the first new feature is the single feature that captures the most variability among the objects in the feature space, the second is the single feature that captures most of the remaining variability, and so on. Thus the first principal component (PC), the first new feature or dimension, is the "best" one-dimensional reflection of the data, the first and second PCs are the "best" two-dimensional reflection, and so on. Figure 22.9*A* shows the percentage of the variance in the perceptual feature space explained by each of the first 10 PCs. As can be seen, the effective dimensionality of the odor profile space was much smaller than 146, with the first two dimensions or PCs accounting for 40.1% of the total variance in the odor profiles and the first four accounting for 54%. Using these four PCs that explained more than half of the variance, we generated a subspace into which we projected the 144 odorants (figure 22.9*B*). In other words, by applying PCA, we reduced the 146-dimensional feature space into a relatively low-dimensional perceptual space.

To test whether the space formed from the odor profile data can be thought of as a perceptual space, we tested whether distances in the space reflected perceptual odor similarity. We pseudo-randomly selected nine odorants that span the space (figure 22.9*B*) and then compared pairwise perceptual similarity between all odorant pairs to pairwise Euclidean distances within the space. Euclidian distance in the space was a powerful predictor of odor perceptual similarity (figure 22.9*C*). In a second experiment, we tested our perceptual space using an implicit similarity task. Using a subset of five odorants from the nine previously used, we presented subjects with a forced choice speeded reaction time task in which subjects were presented with two odorants in succession and required to indicate as quickly as possible whether they were the same or different. In such tasks, the time to respond correctly is expected to be longer for more similar odorants and shorter for more dissimilar odorants (Wise & Cain, 2000; Abraham et al., 2004; Khan & Sobel, 2004; Rinberg, Koulakov, & Gelperin, 2006). As we expected, subjects took longer to make correct judgments for more similar odorant pairs than for dissimilar odorant pairs (figure 22.9*D*), where similarity was derived from our

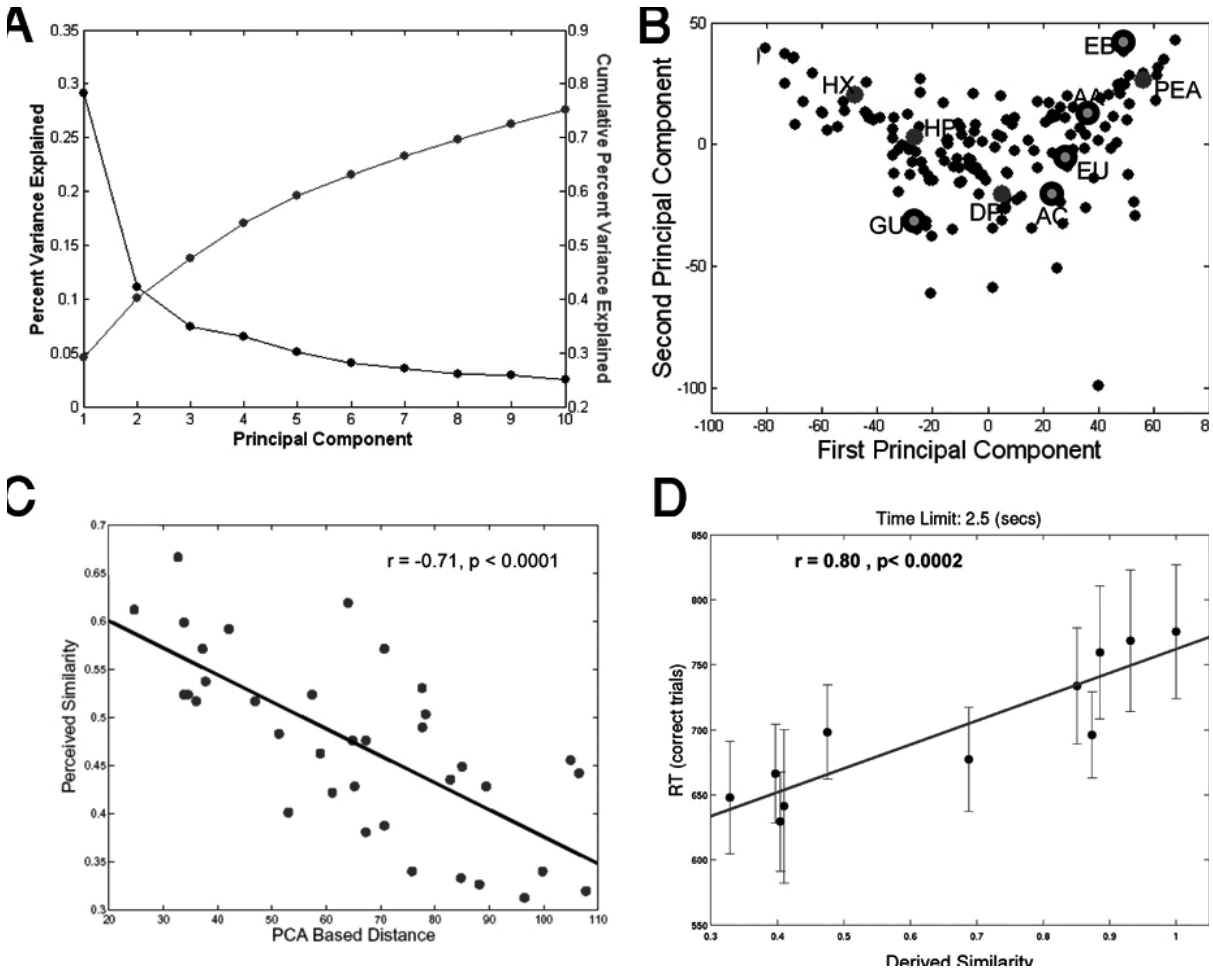
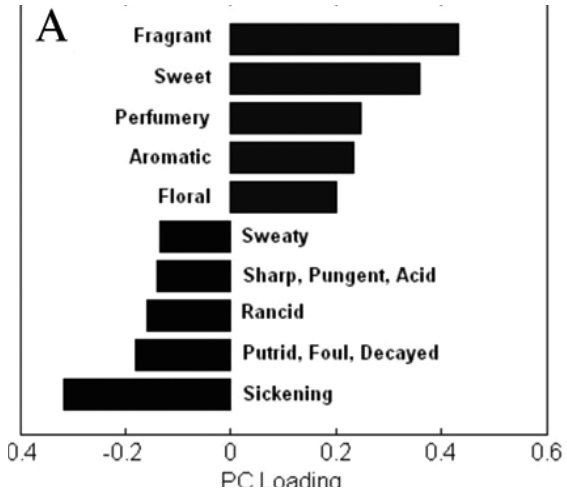


FIGURE 22.9 Olfactory perceptual space. (*A*) The proportion of variance in perceptual descriptions is explained by each of the PCs (starting at 0.3), and the cumulative variance is explained (starting at 0.05). (*B*) The 144 odorants projected into a two-dimensional space made of the first and second PCs. The nine odorants used in experiment 1 (acetophenone (AC), amyl acetate (AA), diphenyl oxide (DP), ethyl butyrate (EB), eugenol (EU), guaiacol (GU), heptanal (HP), hexanoic acid (HX), and phenyl ethanol (PEA) are in enlarged circles, and the five odorants used in experiment 2 (acetophenone, amyl acetate, ethyl butyrate, eugenol, guaiacol) are in further enlarged circles. (*C*) For the nine odorants, the correlation between explicit perceived similarity ratings and PCA-based distance for all pairwise comparisons. Odorants closer in the perceptual space were perceived as more similar. (*D*) Reaction time for correct trials in a forced choice same-different task using five of the nine odorants. Error bars reflect SE. The reaction time was longer for odorant pairs that were closer in PCA-based space, thus providing an implicit validation of the perceptual space.

perceptual space. Thus in both explicit and implicit similarity tasks, our derived perceptual space corresponded to subjects' judgments of similarity. Odorants near each other in our space were perceived as similar, and odorants distant from one another were perceived as dissimilar.

Having validated the space, we set out to identify its principal axis. A first indication to the identity of PC1 is in the descriptors that flank it, that were SWEET, PERFUMERY, AROMATIC, FLORAL, and LIGHT on one end and SICKENING, PUTRID-FOUL-DECAYED, RANCID, SHARP-PUNGENT-ACID, and SWEATY at the other (figure 22.10*A*). An intuitive name for an axis spanning these descriptors is *perceptual pleasantness* (we use the term *pleasantness* for the sake of simplicity, yet we are referring to the continuum from unpleasant to pleasant, also referred to as *perceptual valence* or *hedonic tone*). To test this intuitive label, in a separate experiment, we asked subjects to rank the pleasantness of the odorants, and then compared the difference in pleasantness to distances along PC1. The two measures were strongly correlated (figure 22.10*B*), indicating that our intuitive label of *pleasantness* for PC1 was valid.

Our characterization of the first PC as pleasantness is in agreement with previous research (Richardson & Zucco, 1989). Pleasantness is the primary perceptual aspect humans use to discriminate odorants (Schiffman, 1974; Godinot & Sicard, 1995) or combine them into groups (Berglund, Berglund, Engen, & Ekman, 1973; Schiffman, Robinson, & Erickson, 1977). Pleasant and unpleasant odorants are



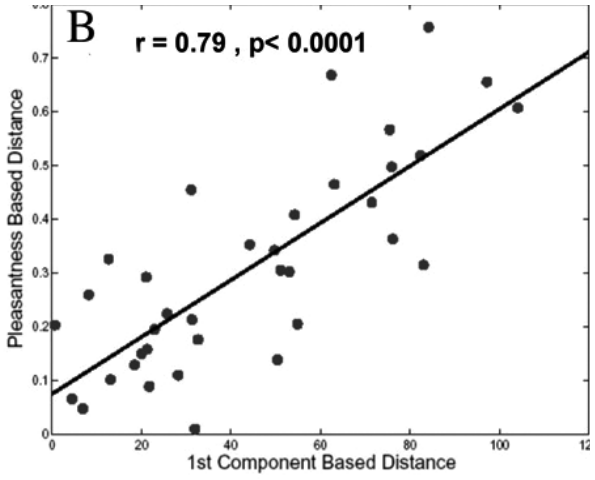


FIGURE 22.10 Identifying pleasantness as the first PC of perception. (*A*) The five descriptors that flanked each end of PC1 of perception. We should stress that here we show the five extreme descriptors only to help give a sense of the PC. This does not reflect a cutoff in any stage of the analysis, but only an esthetic cutoff for the figure. (*B*) For the nine odorants, the correlation between the pairwise difference in pleasantness and the pairwise distance along the first PC. Distance along the first PC was a strong predictor of difference in pleasantness.

evaluated at different speeds (Bensafi et al., 2002) and by dissociable neural substrates, as evidenced in both electrophysiological recordings (Kobal, Hummel, & Vantoller, 1992; AlaouiIsmaili, Rubin, Rada, Dittmar, & Vernet-Maury, 1997; Pause & Krauel, 2000; Masago, Shimomura, Iwanaga, & Katsuura, 2001) and functional neuroimaging studies (Zald & Pardo, 1997; Royet et al., 2000; Gottfried, Deichmann, Winston, & Dolan, 2002; Anderson et al., 2003; Rolls, Kringelbach, & de Araujo, 2003; Grabenhorst, Rolls, Margot, da Silva, & Velazco, 2007). Finally, studies with newborns suggest that at least some aspects of olfactory pleasantness may be innate (Steiner, 1979; Soussignan Schaal, Marlier, & Jiang, 1997). Thus our findings are consistent with the view that "it is clearly the hedonic meaning of odor that dominates odor perception" (Engen, 1982).

Building a physicochemical molecular descriptor space

Having identified the primary perceptual axis of olfaction, we set out to ask whether any physicochemical dimension may be linked with it. Using structural chemistry software (Dragon), we obtained 1514 physicochemical descriptors for each type of odorant. These descriptors were of many types, for example, atom counts, functional group counts, counts of types of bonds, molecular weights, topological descriptors, and so on. We then used the same PCA procedure to reduce the dimensionality of the physicochemical space. Applying PCA revealed that the effective dimensionality of the space of descriptors was much lower than the apparent dimensionality of 1514. Figure 22.11*A* shows the percent variance explained by each of the first 10 PCs. The first PC accounted for approximately 32% of the variance, and the first 10 accounted for approximately 70% of the variance. Figure 22.11*B* shows the five descriptors that anchored the first PC of the space. The full names of these physicochemical descriptors are listed in the legend of figure 22.11.

Characterizing the primary dimensions of the PCA space of the physicochemical descriptors is more challenging than the task for the perceptual space, because of both the set size and the variety of descriptors involved. Nevertheless, giving them a coherent general character is possible and useful for providing a sense of what information they might capture. The first physicochemical PC was weighted at one end by factors that are reasonable proxies for molecular size or weight: the sum of the atomic van der Waals volumes is essentially a crude count of atoms, as is the count of the number of nonhydrogen atoms and the self-returning walk count of order one for nonhydrogen atoms (which is actually identical to a count of the nonhydrogen atoms). The characterization of these descriptors as indices of "weight" is borne out by a very high weighting that "molecular weight" itself has on this side of the first PC.

At the other end of the dimension are a series of topological descriptors that vary with the "extent" of a molecule. In fact, all five of the descriptors are average eigenvectors of distance or adjacency matrices, normalized in slightly different ways: average eigenvector coefficient sum from electronegativity-weighted distance matrix, and average eigenvector coefficient sum from Z-weighted distance matrix (Barysz matrix), average eigenvector coefficient sum from mass-weighted distance matrix, and average eigenvector coefficient sum from distance matrix, average eigenvector coefficient sum from adjacency matrix. Each of these measures increases as the denseness of the atomic connections

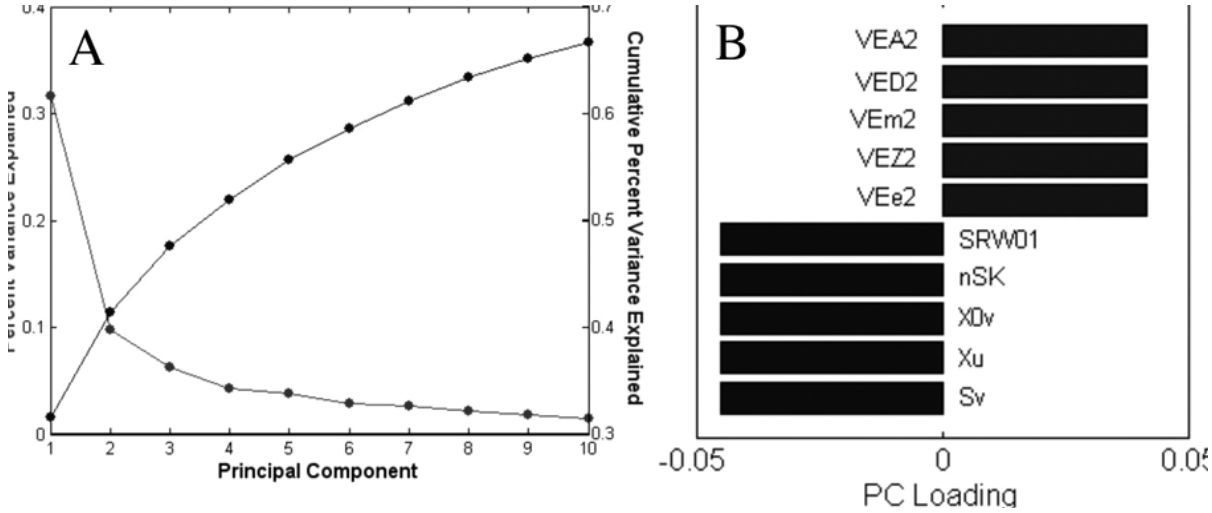


FIGURE 22.11 Reducing dimensionality of physicochemical space. (*A*) The proportion of variance in physicochemical descriptors is explained by each of the PCs (starting at 0.32), and the cumulative variance is explained (starting at 0.01). (*B*) The five descriptors that weighted most heavily at the ends of PC1 of physicochemical space. The descriptors are as follows: Sv, sum of atomic van der Waals volumes (scaled on carbon atom); Xu, Xu index; X0v, pleasantness connectivity index; nSK, number of non-H atoms; SRW01, self-returning walk count of order 01 (number of non-H atoms, nSK); VEe2, average eigenvector coefficient sum from electronegativity weighted distance matrix; VEZ2, average eignvector coefficient sum from *z*-weighted distance matrix (Barysz matrix); Vem2, average eigenvector coefficient sum from mass weighted distance matrix; VEA2, average eigenvector coefficient sum from adjacency matrix; VED2, average eigenvector coefficient sum from distance matrix.

increases, that is, as the number of atoms is packed more closely together. In combination, then, these two extremes anchor a dimension that characterizes the amount and distribution of mass within a molecule. Thus the first PC can be thought of as a measure of the "compactness" of a molecule.

Building a model from physical to perceptual space

We used the same procedure to construct two spaces: a perceptual space that was derived from an initially high-dimensional odor descriptor space and a physicochemical space that was derived from an initially high-dimensional molecular descriptor space. Both spaces were constructed by using PCA, which generates ordered sets of orthogonal axes, constructed to maximize the variance they capture in the original feature space. Because the axes are orthogonal by construction, that is, uncorrelated using a linear Pearson correlation statistic, we can compare them independently.

For each of the first four perceptual PCs, we asked whether they were correlated with any of the first seven physicochemical PCs. Strikingly, we noted that the strongest correlation that we observed was between the first perceptual PC and the first physicochemical PC (figure 22.12*A*). In other words, the single optimal axis for explaining the variance in the physicochemical data was the best predictor of the single optimal axis for explaining the variance in the perceptual data. That the most important dimension in olfactory perception should be the best correlate of the most discriminating physicochemical measures suggests that, as with other senses, the olfactory system has evolved to exploit a fundamental regularity in the physical world.

Having established that the physicochemical space maps onto the perceptual space, we next built linear predictive models through a cross-validation procedure. We then split the Dravnieks data in half, modeled one half, and used this to predict PC1 of the other half. We repeated this 1000 times and obtained a modest but significant prediction of PC1 (odorant pleasantness) based on physicochemical attributes (figure 22.12*B*). To test the generality of this finding, we predicted the pleasantness of 104 odorants that we had never smelled before and that were not used by Dravnieks or us at any stage. We then obtained these odorants and collected pleasantness estimates from three culturally diverse groups of subjects (Americans in Berkeley, California, in the United States; Israeli Jews in Rehovot in Israel; and rural Israeli Muslim Arabs in the village of Dir El Asad in the Northern Galali in Israel), each tested with more than 20 odorants. In each case, we obtained a similarly accurate and significant prediction of odorant pleasantness based on odorant structure (figure 22.13).

Using odor space to predict neural activity in the olfactory system

PC1 of physicochemical structure, or compactness, is a single axis. However, it is multidimensional in the sense

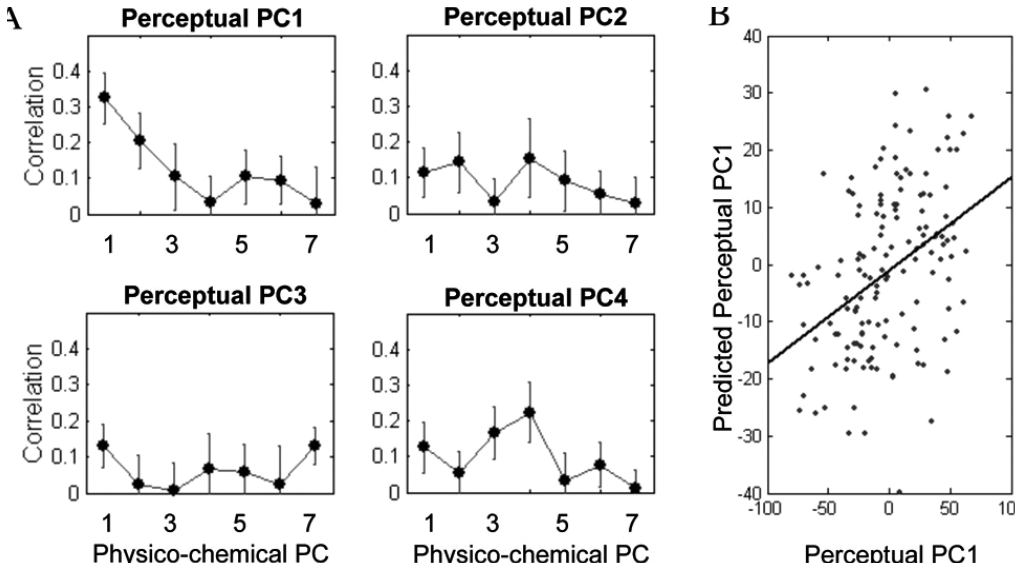


FIGURE 22.12 Relating physicochemical space to perceptual space. (*A*) The correlation between the first to fourth (descending in the figure) perceptual PC and each of the first seven physicochemical PCs for the 144 odorants. Error bars reflect the SE from 1000 bootstrap replicates. The best correlation was between the first PC of perception and the first PC of physicochemical space. This correlation was significantly larger than all other correlations. (*B*) For the 144 odorants, the correlation between their actual first perceptual PC value and the value our model predicted from their physicochemical data ($r = 0.59$; $F(1, 136) = 10.62$; $p = 0.0001$).

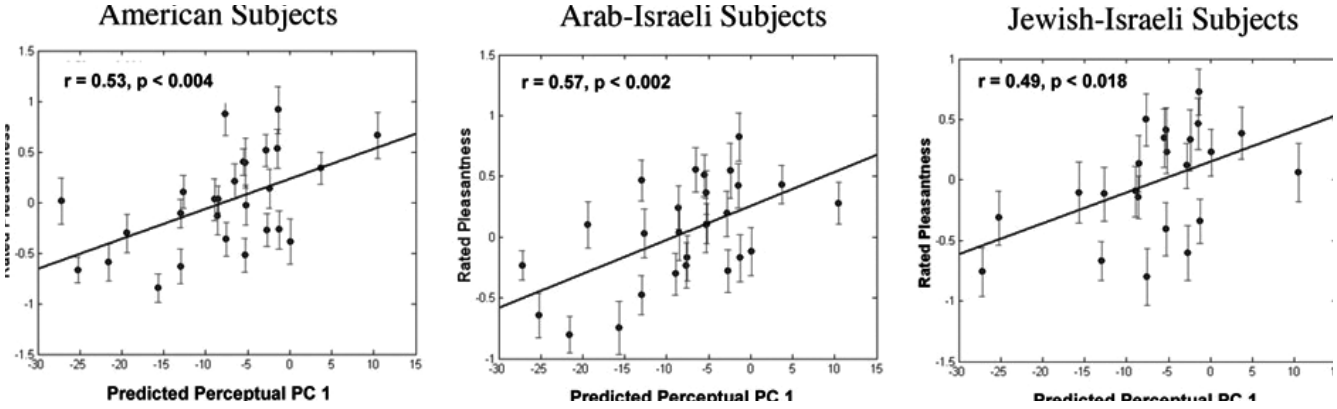


FIGURE 22.13 Cross-cultural validation. Twenty-seven odorous molecules not commonly used in olfactory studies and not previously tested by us were presented to three cultural groups of naïve subjects: Americans (23 subjects), Arab Israelis (22 subjects), and Jewish Israelis (20 subjects). In all cases, our predictions of odorant pleasantness were similar and significant.

that more than 1500 known features contributed to it with known weights. In other words, we can represent each odorant as a single value reflecting its compactness (its PC1 score), or we can represent each odorant as a vector of more than 1500 values (although PC1 was generated with 1514 values, Dragon software will generate up to 1664 values). When using the former approach, the distance between two odorants is the difference in PC1 values. When using the latter approach, one can compute the distance between any two odorants by the square root of the sum of squares of the differences between descriptors.

Given that we have generated an olfactory metric, we can reanalyze previously collected data by reordering the studied odorants using the above described vector-type representation. With this in mind, Rafi Haddad and colleagues in our lab revisited nine previously published data sets as well as one novel data set for which we knew the odorants used but did not know the neural response (Haddad et al., 2008). We found that our novel metric was always better at accounting for neural responses than the specific metric used in each study (e.g., carbon chain length). Moreover, this single metric was applicable across studies that used different olfactory neurons, different model systems, and different neuronal

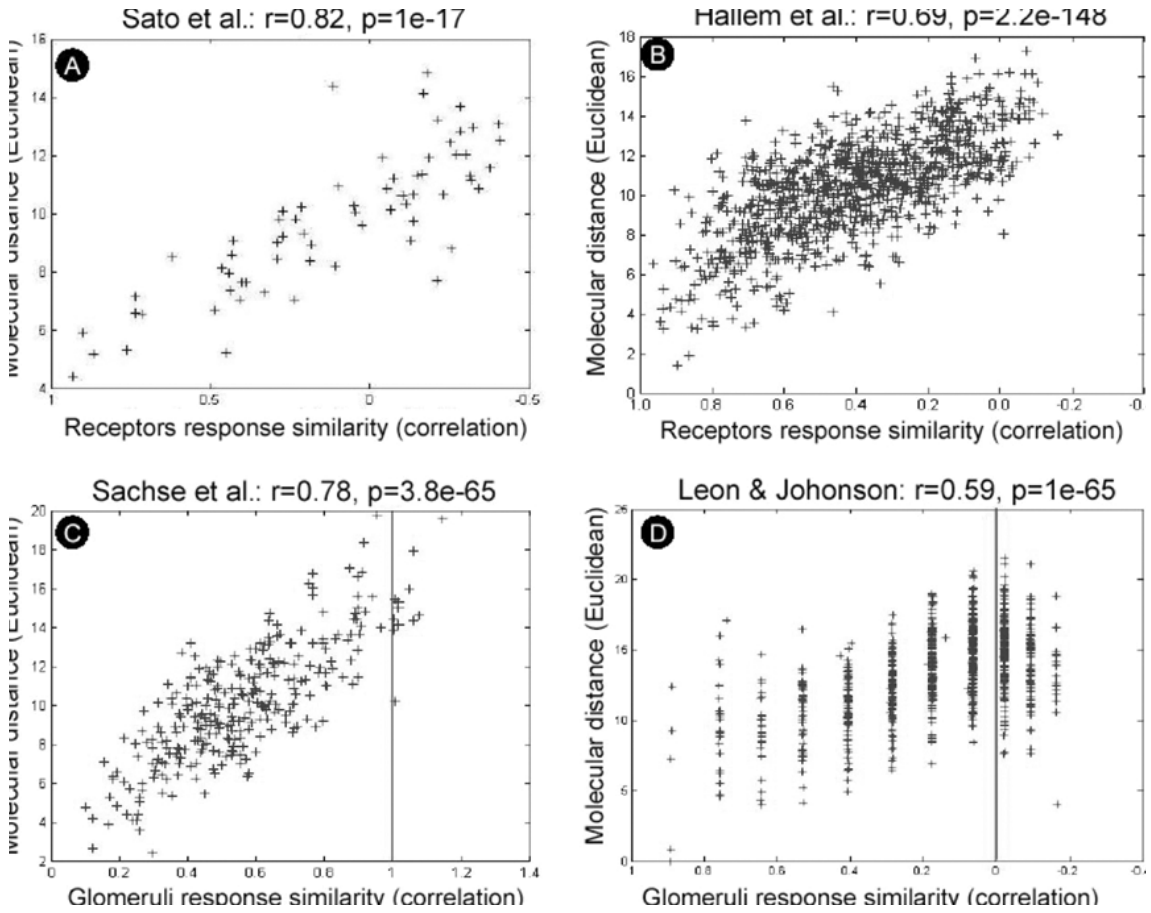


FIGURE 22.14 Correlation plots for unrelated data sets. The graphs demonstrate the ability of the multidimensional olfactory metric to predict neural activity in the olfactory system. Graphs *A* and *B* are from data set reporting RN responses (the similarity was the measured Pearson correlation and thus can range from 1 to −1). Graphs *C* and *D* are from data sets reporting GLO responses. In the *C* and *D* data sets, the *r* value was positive as long as the response pattern similarity was between 0 to 1, and it was either negative or low when the response pattern similarity was negative (the right part of the red line). (*A*) Hallem et al. data set. (*B*) Sato et al. data set. (*C*) Sachse et al. data set (*D*) Leon and Johonson data set. (See color plate 27.)

response measurement techniques and odorants varying along different feature types. In other words, our approach enabled us to use odorant structure to predict olfactory perception in human subjects (Khan et al., 2007) and odor-induced neural activity in nonhuman animals (Haddad et al., 2008) (figure 22.14).

Conclusions

Although the neuroanatomy of the olfactory system is well described and the molecular mechanisms of olfactory transduction are well understood, overall coding of olfaction remains a mystery, in the sense previously noted whereby an olfactory percept cannot be predicted from an olfactory stimulus structure. Our recent efforts have, in our view, made a step in this direction. However, this remains an initial step in what is a long path. To reiterate, we first reduced the dimensionality of olfactory perception and joined others (Engen, 1982) in observing that pleasantness is the principal perceptual aspect of olfaction. We next reduced the dimensionality of physicochemical properties, and identified a primary axis of physicochemical structure. We found that 144 molecules were similarly ordered by these two independently obtained principal axes: one for perception and one for physicochemical structure. In other words, when measures useful to chemists with no a priori connection to any particular percepts were analyzed, those physicochemical measures that were best at discriminating a set of molecules were found to be precisely those that were most correlated with the perception of olfactory pleasantness. It is in identification of this privileged link that we add to the work of Schiffman, Amoore, Dravnieks, and others, who together laid the groundwork for this approach between the early 1950s and the late 1970s (Amoore, 1963; Laffort & Dravnieks, 1973; Schiffman, 1974).

As this link between physicochemical organization and perceptual organization is a key concept in this chapter, we think that it worth reiterating more plainly: Imagine that you were given 1514 physicochemical descriptors (but no perceptual descriptors) for each of 144 molecules, about which you were told nothing else, not even their names, and were instructed to order these descriptors on a single axis that best captures the variance observed in the full set of molecular descriptors. Now imagine that having produced such an axis, you were to arrange all of your 144 odorants on it, those with the lowest scores at one end and those with the highest scores at the other. Now imagine that you walk down your line and smell the odorants in turn. What we have found is that the odorants at one end of your line will smell relatively unpleasant and those at the other will smell relatively pleasant.

That perceptual pleasantness is a reflection of optimal physicochemical discrimination might at first appear at odds with the notion that olfactory pleasantness is variable across both individuals and cultures (Moncrieff, 1966; Pangborn, 1975; Wysocki, Pierce, & Gilbert, 1991; Ayabe-Kanamura et al., 1998), as well as malleable within individuals over time (Cain & Johnson, 1978; Hudson, 1999; Stevenson & Repacholi, 2003). In our own cross-cultural experiments, we found that judgments of pleasantness were less culturally variable than one might expect. We propose that the notion of large cultural variability in pleasantness may be overstated, owing to the salience of exceptions to the rule. Those cases in which one person finds an odorant pleasant and another finds the same odorant unpleasant are in such contrast to our usual experience ("How in the world can you find *THAT* smell to be pleasant?"), that they attain increased salience in our view of olfaction. Whereas it is true that these exceptions exist, there is nevertheless overwhelming agreement on gross olfactory pleasantness. Individuals may differ in the extent to which they find one perfume more pleasant than another or one rotting substance more disagreeable than another, but rarely will they find the rotting odor decidedly pleasant or the perfume decidedly unpleasant.

This is not to say that all of olfactory perception is common or innate. As in other senses, the perception of odor and of pleasantness is a complex process involving both innately tuned and learned components. The olfactory system is known for plasticity at multiple levels (Graziadei et al., 1979; Wilson, Best, & Sullivan, 2004; Barkai, 2005; Mandairon, Stack, Kiselycznyk, & Linster, 2006) and is characterized by experience-dependent learning (Keverne, 1995; Wilson, 2003; Wilson & Stevenson, 2003; Davis, 2004; Wilson & Stevenson, 2006) that reflects an advantageous evolutionary mechanism. Olfactory perception and subsequent neural representations are significantly influenced by several aspects that are clearly unrelated to physicochemical structure, such as context (Schoenbaum & Eichenbaum, 1995; Kay & Laurent, 1999; Herz & von Clef, 2001; Djordjevic et al., 2008), expectation (de Araujo, Rolls, Velazco, Margot, & Cayeux, 2005; Zelano et al., 2005), multisensory convergence (Haberly, 2001; Gottfried & Dolan, 2003; Rolls, 2004), conditioning (Stevenson, Borakes, & Prescott, 1998; Li, Howard, Parrish, & Gottfried, 2008), and various top-down state-dependent modulatory influences (Pager, 1983; Critchley & Rolls, 1996; Kay & Freeman, 1998; Murakami, Kashiwadani, Kirino, & Mori, 2005). For example, an unfamiliar odor will smell sweeter after pairing with sweet taste (Stevenson et al., 1998), a cherry odor may smell smoky following previous association with a smoky odor (Stevenson, 2001a, 2001b), and the same test odor will smell more pleasant when labeled "cheddar cheese" than when labeled "body odor" (de Araujo et al., 2005). Experience-dependent neural representations of odor quality may be rapidly updated through perceptual experience, such as prolonged exposure to a target odorant (Li, Luxenberg, Parrish & Gottfried, 2006), aversive conditioning (Li et al., 2008), or congruent semantic information (Gottfried & Dolan, 2003).

The dynamics of this plastic aspect of olfactory perception will obviously not be reflected in physicochemical structure. However, a portion of olfactory perception is innate and hard-wired (Blanchard et al., 1989; Zangrossi & File, 1992; Dielenberg, Hunt, & McGregor, 2001; Dielenberg & McGregor, 2001), and our results concern this portion. Innately, the olfactory system evolved to maximize discrimination between molecules, and as with other senses, its perceptual organization has evolved to reflect the axis of maximal variance in stimulus space. Thus in perceiving pleasantness, the olfactory system may be understood to be a programmed widely tuned discriminator of molecules rather than a haphazard or arbitrary translator of molecular properties. In other words, what is on one hand surprising is on the other hand inevitable.

REFERENCES

Abraham, N. M., Spors, H., Carleton, A., Margrie, T. W., Kuner, T., & Schaefer, A. T. (2004). Maintaining accuracy at the expense of speed: Stimulus similarity defines odor discrimination time in mice. *Neuron*, *44*, 865–876.

Adrian, E. D. (1942). Olfactory reactions in the brain of the hedgehog. *J. Physiol.*, *100*, 459–473.

AlaouiIsmaili, O., Robin, O., Rada, H., Dittmar, A., & VernetMaury, E. (1997). Basic emotions evoked by odorants: Comparison between autonomic responses and self-evaluation. *Physiol. Behav.*, *62*, 713–720.

Allison, A. (1954). The secondary olfactory areas in the human brain. *J. Anat.*, *88*, 481–488.

Amoore, J. E. (1963). Stereochemical theory of olfaction. *Nature*, *198*, 271–272.

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., Gabrieli, J. D., & Sobel, N. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nat. Neurosci.*, *6*, 196–202.

Axel, R. (1995). The molecular logic of smell. *Sci. Am.*, *273*, 154–159.

Ayabe-Kanamura, S., Schicker, I., Laska, M., Hudson, R., Distel, H., Kobayakawa, T., & Saito, S. (1998). Differences in perception of everyday odors: A Japanese-German cross-cultural study. *Chem. Senses*, *23*, 31–38.

Bakalyar, H. A., & Reed, R. R. (1990). Identification of a specialized adenylyl cyclase that may mediate odorant detection. *Science*, *250*, 1403–1406.

Barkai, E. (2005). Dynamics of learning-induced cellular modifications in the cortex. *Biol. Cybern.*, *92*, 360–366.

Bathellier, B., Buhl, D. L., Accolla, R., & Carleton, A. (2008). Dynamic ensemble odor coding in the mammalian olfactory bulb: Sensory information at different timescales. *Neuron*, *57*, 586–598.

Bedard, A., & Parent, A. (2004). Evidence of newly generated neurons in the human olfactory bulb. *Brain Res. Dev. Brain Res.*, *151*, 159–168.

Bensafi, M., Pierson, A., Rouby, C., Farget, V., Bertrand, B., Vigouroux, M., Jouvent, R., & Holley, A. (2002). Modulation of visual event-related potentials by emotional olfactory stimuli. *Neurophysiol. Clin.*, *32*, 335–342.

Berglund, B., Berglund, U., Engen, T., & Ekman, G. (1973). Multidimensional analysis of 21 odors. *Scand. J. Psychol.*, *14*, 131–137.

Bhalla, U. S., & Bower, J. M. (1997). Multiday recordings from olfactory bulb neurons in awake freely moving rats: Spatially and temporally organized variability in odorant response properties. *J. Comput. Neurosci.*, *4*, 221–256.

Blanchard, J., Blanchard, D. C., & Hori, K. (1989). An ethoexperimental approach to the study of defense. In R. J. Blanchard, P. F. Brain, & D. C. Blanchard (Eds.), *Ethoexperimental approaches to the study of behavior* (pp. 114–136). Dordrecht: Kluwer Academic Publishers.

Boekhoff, I., Tareilus, E., Strotmann, J., & Breer, H. (1990). Rapid activation of alternative second messenger pathways in olfactory cilia from rats by different odorants. *EMBO J.*, *9*, 2453–2458.

Bressler, S. L. (1990). The gamma wave: A cortical information carrier? *Trends Neurosci.*, 13, 161–162.

Bressler, S. L., & Freeman, W. J. (1980). Frequency analysis of olfactory system EEG in cat, rabbit, and rat. *Electroencephalogr. Clin. Neurophysiol.*, *50*, 19–24.

Brunet, L. J., Gold, G. H., & Ngai, J. (1996). General anosmia caused by a targeted disruption of the mouse olfactory cyclic nucleotide-gated cation channel. *Neuron*, *17*, 681–693.

Brunjes, P. C., Illig, K. R., & Meyer, E. A. (2005). A field guide to the anterior olfactory nucleus (cortex). *Brain Res. Brain Res. Rev.*, *50*, 305–335.

Buck, L., & Axel, R. (1991). A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell*, *65*, 175–187.

Buck, L. B. (1996). Information coding in the vertebrate olfactory system. *Annu. Rev. Neurosci.*, *19*, 517–544.

Cain, W. S. (1977). Differential sensitivity for smell: "Noise" at the nose. *Science*, *195*, 796–798.

Cain, W. S., & Johnson, F., Jr. (1978). Lability of odor pleasantness: Influence of mere exposure. *Perception*, *7*, 459–465.

Carmichael, S. T., Clugnet, M. C., & Price, J. L. (1994). Central olfactory connections in the macaque monkey. *J. Comp. Neurol.*, *346*, 403–434.

Carr, V. M., Farbman, A. I., Colletti, L. M., & Morgan, J. I. (1991). Identification of a new nonneuronal cell type in rat olfactory epithelium. *Neuroscience*, *45*, 433–449.

Chess, A., Simon, I., Cedar, H., & Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene-expression. *Cell*, *78*, 823–834.

Cleland, T. A., & Linster, C. (2003). Central olfactory structures. In R. L. Doty (Ed.), *Handbook of olfaction and gustation* (2nd ed., pp. 165–180). New York: Marcel Dekker.

Clerico, D. M., To, W. C., & Lanza, D. C. (2003). Anatomy of the human nasal passages. In R. L. Doty (Ed.), *Handbook of olfaction and human gustation* (2nd ed., pp. 1–16). New York: Marcel Dekker.

Critchley, H. D., & Rolls, E. T. (1996). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J. Neurophysiol.*, *75*, 1673–1686.

Davis, R. L. (2004). Olfactory learning. *Neuron*, *44*, 31–48.

de Araujo, I. E., Rolls, E. T., Velazco, M. I., Margot, C., & Cayeux, I. (2005). Cognitive modulation of olfactory processing. *Neuron*, *46*, 671–679.

de Olmos, J., Hardy, H., & Heimer, L. (1978). The afferent connections of the main and the accessory olfactory bulb formations in the rat: An experimental HRP-study. *J. Comp. Neurol.*, *15*, 213–244.

Di Prisco, G., & Freeman, W. (1985). Odor-related bulbar EEG spatial pattern analysis during appetitive conditioning in rabbits. *Behav. Neurosci.*, *99*, 964–978.

Dielenberg, R. A., Hunt, G. E., & McGregor, I. S. (2001). "When a rat smells a cat": The distribution of Fos immunoreactivity in rat brain following exposure to a predatory odor. *Neuroscience*, *104*, 1085–1097.

Dielenberg, R. A., & McGregor, I. S. (2001). Defensive behavior in rats towards predatory odors: A review. *Neurosci. Biobehav. Rev.*, *25*, 597–609.

Djordjevic, J., Lundstrom, J. N., Clement, F., Boyle, J. A., Pouliot, S., & Jones-Gotman, M. (2008). A rose by any other name: Would it smell as sweet? *J. Neurophysiol.*, *99*, 386–393.

Doty, R. L. (1995). Intranasal trigeminal chemoreception: Anatomy, physiology, and psychophysics. In R. L. Doty (Ed.), *Handbook of olfaction and gustation* (pp. 821–834). New York: Marcel Dekker.

Dravnieks, A. (1982). Odor quality: Semantically generated multidimensional profiles are stable. *Science*, *218*, 799–801.

Dravnieks, A. (1985). *Atlas of odor character profiles*. West Conshohocken, PA: ASTM Press.

Dulac, C. (2000). Sensory coding of pheromone signals in mammals. *Curr. Opin. Neurobiol.*, *10*, 511–518.

Edwards, D. A., Mather, R. A., & Dodd, G. H. (1988). Spatial variation in response to odorants on the rat olfactory epithelium. *Experientia*, *44*, 208–211.

Eeckman, F., & Freeman, W. (1990). Correlations between unit firing and EEG in the rat olfactory system. *Brain Res.*, *528*, 238–244.

Engen, T. (1982). *The perception of odors*. New York: Academic Press.

Firestein, S. (2001). How the olfactory system makes sense of scents. *Nature*, *413*, 211–218.

Firestein, S., Picco, C., & Menini, A. (1993). The relation between stimulus and response in olfactory receptor-cells of the tiger salamander. *J. Physiol. Lond.*, *468*, 1–10.

Firestein, S., & Werblin, F. (1989). Odor-induced membrane currents in vertebrate-olfactory receptor neurons. *Science*, *244*, 79–82.

Freeman, W. J. (1960). Correlations of electrical activity of prepyriform cortex and behavior in cat. *J. Neurophys.*, *23*, 111–131.

Freeman, W. J., & Barrie, J. (1994). Chaotic oscillations and the genesis of meaning in cerebral cortex. In G. Buzsaki, R. Llinas, W. Singer, A. Berthoz, & C. Y. Berlin (Eds.), *Temporal coding in the brain* (pp. 13–37). Berlin: Springer-Verlag.

Friedrich, R. W. (2002). Real time odor representations. *Trends Neurosci.*, *25*, 487–489.

Friedrich, R. W., & Laurent, G. (2001). Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. *Science*, *291*, 889–894.

Getchell, M. L., & Getchell, T. V. (1991). Immunohistochemical localization of components of the immune barrier in the olfactory mucosae of salamanders and rats. *Anat. Rec.*, *231*, 358–374.

Gibbons, B. (1986). The intimate sense of smell. *National Geographic*, *170*, 324–361.

Godinot, N., & Sicard, G. (1995). Odor categorization by human-subjects: An experimental approach. *Chem. Senses*, *20*, 101.

Goldman, A. L., Van der Goes van Naters, W., Lessing, D., Warr, C. G., & Carlson, J. R. (2005). Coexpression of two functional odor receptors in one neuron. *Neuron*, *45*, 661–666.

Gottfried, J. A., Deichmann, R., Winston, J. S., & Dolan, R. J. (2002). Functional heterogeneity in human olfactory cortex: An event-related functional magnetic resonance imaging study. *J. Neurosci.*, *22*, 10819–10828.

Gottfried, J. A., & Dolan, R. J. (2003). The nose smells what the eye sees: Crossmodal visual facilitation of human olfactory perception. *Neuron*, *39*, 375–386.

Gottfried, J. A., Winston, J. S., & Dolan, R. J. (2006). Dissociable codes of odor quality and odorant structure in human piriform cortex. *Neuron*, *49*, 467–479.

Grabenhorst, F., Rolls, E. T., Margot, C., da Silva, M. A., & Velazco, M. I. (2007). How pleasant and unpleasant stimuli combine in different brain regions: Odor mixtures. *J. Neurosci.*, *27*, 13532–13540.

Gray, C. M., & Skinner, J. E. (1988). Centrifugal regulation of neuronal activity in the olfactory bulb of the waking rabbit as revealed by reversible cryogenic blockade. *Exp. Brain Res.*, *69*, 378–386.

Graziadei, P. P., Levine, R. R., & Monti Graziadei, G. A. (1979). Plasticity of connections of the olfactory sensory neuron: Regeneration into the forebrain following bulbectomy in the neonatal mouse. *Neuroscience*, *4*, 713–727.

Graziadei, P. P., & Monti, Graziadei, A. G. (1983). Regeneration in the olfactory system of vertebrates. *Am. J. Otolaryngol.*, *4*, 228–233.

Greer, C. A., Stewart, W. B., Kauer, J. S., & Shepherd, G. M. (1981). Topographical and laminar localization of 2-deoxyglucose uptake in rat olfactory-bulb induced by electrical-stimulation of olfactory nerves. *Brain Res.*, *217*, 279–293.

Guthrie, K. M., Anderson, A. J., Leon, M., & Gall, C. (1993). Odor-induced increases in c-fos mRNA expression reveal an anatomical "unit" for odor processing in olfactory bulb. *Proc. Natl. Acad. Sci. USA*, *90*, 3329–3333.

Guthrie, K. M., & Gall, C. M. (1995). Functional mapping of odor-activated neurons in the olfactory bulb. *Chem. Senses*, *20*, 271–282.

Haberly, L. B. (1985). Neuronal circuitry in olfactory cortex: Anatomy and functional implications. *Chem. Senses*, *10*, 219–238.

Haberly, L. B. (1998). In G. M. Shepherd (Ed.), *The synaptic organization of the brain*. New York: Oxford University Press.

Haberly, L. B. (2001). Parallel-distributed processing in olfactory cortex: New insights from morphological and physiological analysis of neuronal circuitry. *Chem. Senses*, *26*, 551–576.

Haddad, R., Khan, R., Takahashi, Y. K., Mori, K., Harel, D., & Sobel, N. (2008). A metric for odorant comparison. *Nature Methods*, *5*, 425–429.

Hallem, E. A., & Carlson, J. R. (2006). Coding of odors by a receptor repertoire. *Cell*, *125*, 143–160.

Halpern, M. (1987). The organization and function of the vomeronasal system. *Annu. Rev. Neurosci.*, *10*, 325–362.

Harrison, T., & Scott, J. (1986). Olfactory bulb responses to odor stimulation: Analysis of response pattern and intensity relationships. *J. Neurophysiol.*, *56*, 1571–1589.

Herz, R. S., & von Clef, J. (2001). The influence of verbal labeling on the perception of odors: Evidence for olfactory illusions? *Perception*, *30*, 381–391.

Hinds, J. W., Hinds, P. L., & Mcnelly, N. A. (1984). An autoradiographic study of the mouse olfactory epithelium: Evidence for long-lived receptors. *Anat. Rec.*, *210*, 375–383.

Hopfield, J. J. (1995). Pattern-recognition computation using action-potential timing for stimulus representation. *Nature*, *376*, 33–36.

Hopfield, J. J., & Brody, C. D. (2001). What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proc. Natl. Acad. Sci. USA*, *98*, 1282–1287.

Hornung, D. E., & Enns, M. P. (1986). Possible mechanisms for the processes of referred taste and retronasal olfaction. *Chem. Senses*, *11*, 616.

Huard, J. M. T., Youngentob, S. L., Goldstein, B. J., Luskin, M. B., & Schwob, J. E. (1998). Adult olfactory epithelium contains multipotent progenitors that give rise to neurons and non-neural cells. *J. Comp. Neurol.*, *400*, 469–486.

Hudson, R. (1999). From molecule to mind: The role of experience in shaping olfactory function. *J. Comp. Physiol.* [*A*] *185*, 297–304.

Hudson, R., & Distel, H. (1987). Regional autonomy in the peripheral processing of odor signals in newborn rabbits. *Brain Res.*, *22*, 85–94.

Hughes, J., Hendrix, D., Wetzel, N., & Johnston, J. (1969). Correlations between electrophysiological activity from the human olfactory bulb and the subjective response to odoriferous stimuli. In C. Pfaffmann (Ed.), *Olfaction and taste* (3rd ed., pp. 172–191). New York: Academic Press.

Hummel, T. (2000). Assessment of intranasal trigeminal function. *Int. J. Psychophysiol*, *36*, 147–155.

Hummel, T., & Livermore, A. (2002). Intranasal chemosensory function of the trigeminal nerve and aspects of its relation to olfaction. *Int. Arch. Occup. Environ. Health*, *75*, 305–313.

Inaki, K., Takahashi, Y., Nagayama, S., & Mori, K. (2002). Molecular-feature domains with posterodorsal-anteroventral polarity in the symmetrical sensory maps of the mouse olfactory bulb: Mapping of odourant-induced Zif268 expression. *Eur. J. Neurosci.*, *15*, 1563–1574.

Ivic, L., Zhang, C., Zhang, X., Yoon, S. O., & Firestein, S. (2002). Intracellular trafficking of a tagged and functional mammalian olfactory receptor. *J. Neurobiol.*, *50*, 56–68.

Johnson, B. A., Ho, S. L., Xu, Z., Yihan, J. S., Yip, S., Hingco, E. E., & Leon, M. (2002). Functional mapping of the rat olfactory bulb using diverse odorants reveals modular responses to functional groups and hydrocarbon structural features. *J. Comp. Neurol.*, *449*, 180–194.

Johnson, B. A., & Leon, M. (2000a). Odorant molecular length: One aspect of the olfactory code. *J. Comp. Neurol.*, *426*, 330–338.

Johnson, B. A., & Leon, M. (2000b). Modular representations of odorants in the glomerular layer of the rat olfactory bulb and the effects of stimulus concentration. *J. Comp. Neurol.*, *422*, 496–509.

Johnson, B. A., & Leon, M. (2007). Chemotopic odorant coding in a mammalian olfactory system. *J. Comp. Neurol.*, *503*, 1–34.

Johnson, B. A., Woo, C. C., Duong, H. C., Nguyen, V., & Leon, M. (1995). A learned odor evokes an enhanced fos-like glomerular response in the olfactory-bulb of young rats. *Brain Res.*, *699*, 192–200.

Johnson, B. A., Woo, C. C., Hingco, E. E., Pham, K. L., & Leon, M. (1999). Multidimensional chemotopic responses to n-aliphatic acid odorants in the rat olfactory bulb. *J. Comp. Neurol.*, *409*, 529–548.

Johnson, B. A., Woo, C. C., & Leon, M. (1998). Spatial coding of odorant features in the glomerular layer of the rat olfactory bulb. *J. Comp. Neurol.*, *393*, 457–471.

Jones, D. T., & Reed, R. R. (1989). Golf: An olfactory neuron specific-G protein involved in odorant signal transduction. *Science*, *244*, 790–795.

Jourdan, F., Duveau, A., Astic, L., & Holley, A. (1980). Spatial distribution of [14C]2-deoxyglucose uptake in the olfactory bulbs of rats stimulated with two different odours. *Brain Res.*, *188*, 139–154.

Kadohisa, M., & Wilson, D. A. (2006a). Olfactory cortical adaptation facilitates detection of odors against background. *J. Neurophysiol.*, *95*, 1888–1896.

Kadohisa, M., & Wilson, D. A. (2006b). Separate encoding of identity and similarity of complex familiar odors in piriform cortex. *Proc. Natl. Acad. Sci. USA*, *103*, 15206–15211.

Kauer, J. S. (1974). Response patterns of amphibian olfactory bulb neurones to odour stimulation. *J. Neurophysiol.*, *243*, 695–715.

Kauer, J. S. (1988). Real-time imaging of evoked activity in local circuits of the salamander olfactory bulb. *Nature*, *331*, 166–168.

Kay, L. M., & Freeman, W. J. (1998). Bidirectional processing in the olfactory-limbic axis during olfactory behavior. *Behav. Neurosci.*, *112*, 541–553.

Kay, L. M., & Laurent, G. (1999). Odor- and context-dependent modulation of mitral cell activity in behaving rats. *Nat. Neurosci.*, *2*, 1003–1009.

Kent, P. F., & Mozell, M. M. (1992). The recording of odorant-induced mucosal activity patterns with a voltage-sensitive dye. *J. Neurophysiol.*, *68*, 1804–1819.

Kepecs, A., Uchida, N., & Mainen, Z. F. (2006). The sniff as a unit of olfactory processing. *Chem. Senses*, *31*, 167–179.

Kepecs, A., Uchida, N., & Mainen, Z. F. (2007). Rapid and precise control of sniffing during olfactory discrimination in rats. *J. Neurophysiol.*, *98*, 205–213.

Ketchum, K. L., & Haberly, L. B. (1993). Synaptic events that generate fast oscillations in piriform cortex. *J. Neurosci.*, *13*, 3980–3985.

Keverne, E. B. (1995). Olfactory learning. *Curr. Opin. Neurobiol.*, *5*, 482–488.

Keverne, E. B. (1999). The vomeronasal organ. *Science*, *286*, 716–720.

Khan, R., Luk, C., Flinker, A., Aggarwal, A., Lapid, H., Haddad, R., & Sobel, N. (2007). Predicting odor pleasantness from odorant structure: Pleasantness as a reflection of the physical world. *J. Neurosci.*, *27*, 10015–10023.

Khan, R. M., & Sobel, N. (2004). Neural processing at the speed of smell. *Neuron*, *44*, 744–747.

Kobal, G., Hummel, T., & Vantoller, S. (1992). Differences in human chemosensory evoked-potentials to olfactory and somatosensory chemical stimuli presented to left and right nostrils. *Chem. Senses*, *17*, 233–244.

Kratskin, I. L., & Belluzzi, O. (2003). Anatomy and neurochemistry of the olfactory bulb. In R. L. Doty (Ed.), *Handbook of olfaction and gustation* (2nd ed., pp. 139–164). New York: Marcel Dekker.

Krautwurst, D., Yau, K. W., & Reed, R. R. (1998). Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell*, *95*, 917–926.

Laffort, P., & Dravnieks, A. (1973). An approach to a physicochemical model of olfactory stimulation in vertebrates by single compounds. *J. Theor. Biol.*, *38*, 335–345.

Lancet, D., Greer, C. A., Kauer, J. S., & Shepherd, G. M. (1982). Mapping of odor-related neuronal activity in the olfactory bulb by high-resolution, 2-deoxyglucose autoradiography. *Proc. Natl. Acad. Sci. USA*, *79*, 670–674.

Laska, M., Trolp, S., & Teubner, P. (1999). Odor structure-activity relationships compared in human and nonhuman primates. *Behav. Neurosci.*, *113*, 998–1007.

Laurent, G. (2002). Olfactory network dynamics and the coding of multidimensional signals. *Nat. Rev. Neurosci.*, *11*, 884–895.

Lewis, J. E., & Dahl, A. R. (1995). Olfactory mucosa: Composition, enzymatic localization, and metabolism. In R. L. Doty (Ed.), *Handbook of olfaction and gustation* (pp. 33–52). New York: Marcel Dekker.

Li, W., Howard, J. D., Parrish, T. B., & Gottfried, J. A. (2008). Aversive learning enhances perceptual and cortical discrimination of indiscriminable odor cues. *Science*, *319*, 1842–1845.

Li, W., Luxenberg, E., Parrish, T., & Gottfried, J. A. (2006). Learning to smell the roses: Experience-dependent neural plasticity in human piriform and orbitofrontal cortices. *Neuron*, *52*, 1097–1108.

Linster, C., & Smith, B. H. (1997). A computational model of the response of honey bee antennal lobe circuitry to odor mixtures: Overshadowing, blocking and unblocking can arise from lateral inhibition. *Behav. Brain Res.*, *87*, 1–14.

Litaudon, P., Datiche, F., & Cattarelli, M. (1997). Optical recording of the rat piriform cortex activity. *Prog. Neurobiol.*, *52*, 485–510.

Livermore, A., & Laing, D. G. (1996). Influence of training and experience on the perception of multicomponent odor mixtures. *J. Exp. Psychol. Hum. Percept. Perform.*, *22*, 267–277.

Lu, X., & Slotnick, B. (1998). Olfaction in rats with extensive lesions of the olfactory bulbs: Implications for odor coding. *Neuroscience*, *84*, 849–866.

Luo, L., & Flanagan, J. G. (2007). Development of continuous and discrete neural maps. *Neuron*, *56*, 284–300.

Mackay-Sim, A., & Kittel, P. W. (1991). On the life span of olfactory receptor neurons. *Eur. J. Neurosci.*, *3*, 209–215.

Mackay-Sim, A., Shaman, P., & Moulton, D. G. (1982). Topographic coding of olfactory quality: Odorant-specific patterns of epithelial responsivity in the salamander. *J. Neurophysiol.*, *48*, 584–596.

Mainland, J., & Sobel, N. (2006). The sniff is part of the olfactory percept. *Chem. Senses*, *31*, 181–196.

Malnic, B., Hirono, J., Sato, T., & Buck, L. B. (1999). Combinatorial receptor codes for odors. *Cell*, *96*, 713–723.

Mandairon, N., Stack, C., Kiselycznyk, C., & Linster, C. (2006). Broad activation of the olfactory bulb produces long-lasting

changes in odor perception. *Proc. Natl. Acad. Sci. USA*, *103*, 13543–13548.

Masago, R., Shimomura, Y., Iwanaga, K., & Katsuura, T. (2001). The effects of hedonic properties of odors and attentional modulation on the olfactory event-related potentials. *J. Physiol. Anthropol. Appl. Hum. Sci.*, *20*, 7–13.

Meisami, E., Mikhail, L., Baim, D., & Bhatnagar, K. P. (1998). Human olfactory bulb: Aging of glomeruli and mitral cells and a search for the accessory olfactory bulb. *Ann. NY Acad. Sci.*, *855*, 708–715.

Meister, M., & Bonhoeffer, T. (2001). Tuning and topography in an odor map on the rat olfactory bulb. *J. Neurosci.*, *21*, 1351–1360.

Meredith, M. (1991). Sensory processing in the main and accessory olfactory systems: Comparisons and contrasts. *J. Steroid Biochem. Mol. Biol.*, *39*, 601–614.

Meredith, M., & Moulton, D. (1978). Patterned response to odor in single neurones of goldfish olfactory bulb: Influence of odor quality and other stimulus parameters. *J. Gen. Physiol.*, *71*, 615–643.

Miyamichi, K., Serizawa, S., Kimura, H. M., & Sakano, H. (2005). Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *J. Neurosci.*, *25*, 3586–3592.

Mombaerts, P. (1999). Molecular biology of odorant receptors in vertebrates. *Annu. Rev. Neurosci.*, *22*, 487–509.

Mombaerts, P., Wang, F., Dulac, C., Chao, S. K., Nemes, A., Mendelsohn, M., Edmondson, J., & Axel, R. (1996). Visualizing an olfactory sensory map. *Cell*, *87*, 675–686.

Moncrieff, R. W. (1966). *Odour preferences*. London: Leonard Hill.

Moran, D. T., Rowley, J. C., Jafek, B. W., & Lovell, M. A. (1982). The fine-structure of the olfactory mucosa in man. *J. Neurocytol.*, *11*, 721–746.

Mori, K., & Shepherd, G. M. (1994). Emerging principles of molecular signal processing by mitral/tufted cells in the olfactory bulb. *Semin. Cell Biol.*, *5*, 65–74.

Mori, K., & Yoshihara, Y. (1995). Molecular recognition and olfactory processing in the mammalian olfactory system. *Prog. Neurobiol.*, *45*, 585–619.

Morrison, E. E., & Costanzo, R. M. (1990). Morphology of the human olfactory epithelium. *J. Comp. Neurol.*, *297*, 1–13.

Morrison, E. E., & Costanzo, R. M. (1992). Morphology of olfactory epithelium in humans and other vertebrates. *Microsc. Res. Tech.*, *23*, 49–61.

Moulton, D. G. (1976). Spatial patterning of response to odors in peripheral olfactory system. *Physiol. Rev.*, *56*, 578–593.

Moulton, D. G. (1977). Minimum odorant concentrations detectable by the dog and their implications for olfactory receptor sensitivity. In D. Muller-Schwarze & M. M. Mozell (Eds.), *Chemical senses in vertebrates* (pp. 455–464). New York: Plenum.

Mozell, M. M., & Jagodowicz, M. (1973). Chromatographic separation of odorants by the nose: Retention times measured across in vivo olfactory mucosa. *Science*, *181*, 1247–1249.

Murakami, M., Kashiwadani, H., Kirino, Y., & Mori, K. (2005). State-dependent sensory gating in olfactory cortex. *Neuron*, *46*, 285–296.

Mustaparta, H. (1971). Spatial distribution of receptor-responses to stimulation with different odours. *Acta Physiol. Scand.*, *82*, 154–166.

Nakamura, T., & Gold, G. H. (1987). A cyclic nucleotide-gated conductance in olfactory receptor cilia. *Nature*, *325*, 442–444.

Nef, P., Hermansborgmeyer, I., Artierespin, H., Beasley, L., Dionne, V. E., & Heinemann, S. F. (1992). Spatial pattern of receptor expression in the olfactory epithelium. *Proc. Natl. Acad. Sci. USA*, *89*, 8948–8952.

Ninkovic, J., Mori, T., & Gotz, M. (2007). Distinct modes of neuron addition in adult mouse neurogenesis. *J. Neurosci.*, *27*, 10906–10911.

Ohloff, G. (1986). Chemistry of odor stimuli. *Experientia*, *42*, 271–279.

Pace, U., Hanski, E., Salomon, Y., & Lancet, D. (1985). Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature*, *316*, 255–258.

Pagano, S. F., Impagnatiello, F., Girelli, M., Cova, L., Grioni, E., Onofri, M., Cavallaro, M., Etteri, S., Vitello, F., Giombini, S., Solero, C. L., & Parati, E. A. (2000). Isolation and characterization of neural stem cells from the adult human olfactory bulb. *Stem Cells*, *18*, 295–300.

Pager, J. (1983). Unit responses changing with behavioral outcome in the olfactory bulb of unrestrained rats. *Brain Res.*, *289*, 87–98.

Pangborn, R. M. (1975). Cross-cultural aspects of flavour preference. *Food Technol.*, *29*, 34–36.

Pause, B. M., & Krauel, K. (2000). Chemosensory event-related potentials (CSERP) as a key to the psychology of odors. *Int. J. Psychophysiol.*, *36*, 105–122.

Pelosi, P. (2001). The role of perireceptor events in vertebrate olfaction. *Cell. Mol. Life Sci.*, *58*, 503–509.

Pevsner, J., Trifiletti, R. R., Strittmatter, S. M., Sklar, P. B., & Snyder, S. H. (1985). Purification and characterization of a pyrazine odorant binding-protein. *Chem. Senses*, *10*, 397.

Porter, J., Anand, T., Johnson, B., Khan, R. M., & Sobel, N. (2005). Brain mechanisms for extracting spatial information from smell. *Neuron*, *47*, 581–592.

Porter, J., Craven, B., Khan, R. M., Chang, S. J., Kang, I., Judkewitz, B., Volpe, J., Settles, G., & Sobel, N. (2007). Mechanisms of scent-tracking in humans. *Nat. Neurosci.*, *10*, 27–29.

Porter, J., & Sobel, N. (2005). Human scent tracking. In review.

Price, J. L. (1973). An autoradiographic study of complementary laminar patterns of termination of afferent fibers to the olfactory cortex. *J. Comp. Neurol.*, *150*, 87–108.

Price, J. L. (1987). The central olfactory and accessory olfactory systems. In T. E. Finger & W. L. Silver (Eds), *Neurobiology of taste and smell.* (pp. 179–203). New York: Wiley.

Price, J. L. (1990). Olfactory system. In G. Paxinos (Ed.), *The human nervous system* (pp. 979–1001). San Diego: Elsevier Academic.

Price, J. L., & Sprich, W. W. (1975). Observations on the lateral olfactory tract of the rat. *J. Comp. Neurol.*, *162*, 321–336.

Protopapas, A. D., & Bower, J. M. (2001). Spike coding in pyramidal cells of the piriform cortex of rat. *J. Neurophysiol.*, *86*, 1504–1510.

Rajan, R., Clement, J. P., & Bhalla, U. S. (2006). Rats smell in stereo. *Science*, *311*, 666–670.

Raming, K., Krieger, J., Strotmann, J., Boekhoff, I., Kubick, S., Baumstark, C., & Breer, H. (1993). Cloning and expression of odorant receptors. *Nature*, *361*, 353–356.

Rennaker, R. L., Chen, C. F., Ruyle, A. M., Sloan, A. M., & Wilson, D. A. (2007). Spatial and temporal distribution of odorant-evoked activity in the piriform cortex. *J. Neurosci.*, *27*, 1534–1542.

Ressler, K. J., Sullivan, S. L., & Buck, L. B. (1993). A zonal organization of odorant receptor gene-expression in the olfactory epithelium. *Cell*, *73*, 597–609.

Ressler, K. J., Sullivan, S. L., & Buck, L. B. (1994). Information coding in the olfactory system: Evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell*, *79*, 1245–1255.

Richardson, J. T., & Zucco, G. M. (1989). Cognition and olfaction: A review. *Psychol. Bull.*, *105*, 352–360.

Rinberg, D., Koulakov, A., & Gelperin, A. (2006). Speed-accuracy tradeoff in olfaction. *Neuron*, *3*, 351–358.

Roberts, E. (1986). Alzheimer's disease may begin in the nose and may be caused by aluminosilicates. *Neurobiol. Aging*, *7*, 561–567.

Rolls, E. T. (2004). Convergence of sensory systems in the orbitofrontal cortex in primates and brain design for emotion. *Anat. Rec. A Discov. Mol. Cell Evol. Biol.*, *281*, 1212–1225.

Rolls, E. T., Critchley, H. D., & Treves, A. (1996). Representation of olfactory information in the primate orbitofrontal cortex. *J. Neurophysiol.*, *75*, 1982–1996.

Rolls, E. T., Kringelbach, M. L., & de Araujo, I. E. (2003). Different representations of pleasant and unpleasant odours in the human brain. *Eur. J. Neurosci.*, *18*, 695–703.

Roppolo, D., Ribaud, V., Jungo, V. P., Luscher, C., & Rodriguez, I. (2006). Projection of the Gruneberg ganglion to the mouse olfactory bulb. *Eur. J. Neurosci.*, *23*, 2887–2894.

Royet, J. P., Zald, D., Versace, R., Costes, N., Lavenne, F., Koenig, O., & Gervais, R. (2000). Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: A positron emission tomography study. *J. Neurosci.*, *20*, 7752–7759.

Rubin, B. D., & Katz, L. C. (1999). Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron*, *23*, 499–511.

Sallaz, M., & Jourdan, F. (1996). Odour-induced c-fos expression in the rat olfactory bulb: Involvement of centrifugal afferents. *Brain Res.*, *20*, 66–75.

Sato, T., Hirono, J., Tonoike, M., & Takebayashi, M. (1994). Tuning specificities to aliphatic odorants in mouse olfactory receptor neurons and their local-distribution. *J. Neurophysiol.*, *72*, 2980–2989.

Schiffman, S. S. (1974). Physicochemical correlates of olfactory quality. *Science*, *185*, 112–117.

Schiffman, S., Robinson, D. E., & Erickson, R. P. (1977). Multidimensional-scaling of odorants: Examination of psychological and physiochemical dimensions. *Chem. Senses Flavour*, *2*, 375–390.

Schoenbaum, G., & Eichenbaum, H. (1995). Information coding in the rodent prefrontal cortex: I. Single-neuron activity in orbitofrontal cortex compared with that in pyriform cortex. *J. Neurophysiol.*, *74*, 733–750.

Schoenfeld, T. A., & Cleland, T. A. (2006). Anatomical contributions to odorant sampling and representation in rodents: Zoning in on sniffing behavior. *Chem. Senses*, *31*, 131–144.

Scott, J. W., & Brierley, T. (1999). A functional map in rat olfactory epithelium. *Chem. Senses*, *24*, 679–690.

Scott, J. W., Shannon, D. E., Charpentier, J., Davis, L. M., & Kaplan, C. (1997). Spatially organized response zones in rat olfactory epithelium. *J. Neurophysiol.*, *77*, 1950–1962.

Settles, G. S. (2001). *Schlieren and shadowgraph techniques*. New York: Springer.

Shepherd, G. M. (1972). Synaptic organization of the mammalian olfactory bulb. *Physiol. Rev.*, *52*, 864–917.

Shepherd, G. M. (1985). The olfactory system: The uses of neural space for a non-spatial modality. *Prog. Clin. Biol. Res.*, *176*, 99–114.

Shipley, M. T. (1995). Olfactory system. In G. Paxinos (Ed.), *The rat nervous system* (2nd ed., pp. 899–928). San Diego: Academic Press.

Sicard, G., & Holley, A. (1984). Receptor cell responses to odorants: Similarities and differences among odorants. *Brain Res.*, *292*, 283–296.

Singer, B. H., Kim, S., & Zochowski, M. (2007). Binaral interaction and centrifugal input enhances spatial contrast in olfactory bulb activation. *Eur. J. Neurosci.*, *25*, 576–586.

Slotnick, B., & Berman, E. (1980). Transection of the lateral olfactory tract does not produce anosmia. *Brain Res. Bull.*, *5*, 141–145.

Slotnick, B., & Bodyak, N. (2002). Odor discrimination and odor quality perception in rats with disruption of connections between the olfactory epithelium and olfactory bulbs. *J. Neurosci.*, *15*, 4205–4216.

Slotnick, B., & Risser, J. (1990). Odor memory and odor learning in rats with lesions of the lateral olfactory tract and mediodorsal thalamic nucleus. *Brain Res.*, *529*, 23–29.

Slotnick, B., & Schoonover, F. (1993). Olfactory sensitivity of rats with transection of the lateral olfactory tract. *Brain Res.*, *616*, 132–137.

Small, D. M., & Prescott, J. (2005). Odor/taste integration and the perception of flavor. *Exp. Brain Res.*, *166*, 345–357.

Sobel, N., Johnson, B. N., Mainland, J., & Yousem, D. M. (2003). Functional neuroimaging of human olfaction. In R. L. Doty, (Ed.), *Handbook of olfaction and gustation* (2nd ed., pp. 251–273). New York: Marcel Dekker.

Sobel, N., Khan, R. M., Saltman, A., Sullivan, E. V., & Gabrieli, J. D. (1999). The world smells different to each nostril. *Nature*, *402*, 35.

Sobel, N., Prabhakaran, V., Desmond, J. E., Glover, G. H., Goode, R. L., Sullivan, E. V., & Gabrieli, J. D. (1998). Sniffing and smelling: Separate subsystems in the human olfactory cortex. *Nature*, *392*, 282–286.

Soussignan, R., Schaal, B., Marlier, L., & Jiang, T. (1997). Facial and autonomic responses to biological and artificial olfactory stimuli in human neonates: Re-examining early hedonic discrimination of odors. *Physiol. Behav.*, *62*, 745–758.

Spors, H., & Grinvald, A. (2002). Spatio-temporal dynamics of odor representations in the mammalian olfactory bulb. *Neuron*, *34*, 301–315.

Spors, H., Wachowiak, M., Cohen, L. B., & Friedrich, R. W. (2006). Temporal dynamics and latency patterns of receptor neuron input to the olfactory bulb. *J. Neurosci.*, *26*, 1247–1259.

Steiner, J. E. (1979). Human facial expressions in response to taste and smell stimulation. *Adv. Child Dev. Behav.*, *13*, 257–295.

Stevenson, R. J. (2001a). Associative learning and odor quality perception: How sniffing an odor mixture can alter the smell of its parts. *Learn. Motiv.*, *32*, 154–177.

Stevenson, R. J. (2001b). The acquisition of odour qualities. *Q. J. Exp. Psychol. [A]*, *54*, 561–577.

Stevenson, R. J., Boakes, R. A., & Prescott, J. (1998). Changes in odor sweetness resulting from implicit learning of a simultaneous odor-sweetness association: An example of learned synesthesia. *Learn. Motiv.*, *29*, 113–132.

Stevenson, R. J., & Repacholi, B. M. (2003). Age-related changes in children's hedonic response to male body odor. *Dev. Psychol.*, *39*, 670–679.

Stevenson, R. J., & Wilson, D. A. (2007). Odour perception: An object-recognition approach. *Perception*, *36*, 1821–1833.

Stewart, W. B., Kauer, J. S., & Shepherd, G. M. (1979). Functional organization of rat olfactory bulb analysed by the 2-deoxyglucose method. *J. Comp. Neurol.*, *185*, 715–734.

Storan, M. J., & Key, B. (2006). Septal organ of Gruneberg is part of the olfactory system. *J. Comp. Neurol.*, *494*, 834–844.

Strotmann, J., Conzelmann, S., Beck, A., Feinstein, P., Breer, H., & Mombaerts, P. (2000). Local permutations in the glomerular array of the mouse olfactory bulb. *J. Neurosci.*, *20*, 6927–6938.

Strotmann, J., Wanner, I., Helfrich, T., Beck, A., Meinken, C., Kubick, S., & Breer, H. (1994). Olfactory neurons expressing distinct odorant receptor subtypes are spatially segregated in the nasal neuroepithelium. *Cell Tissue Res.*, *276*, 429–438.

Strotmann, J., Wanner, I., Krieger, J., Raming, K., & Breer, H. (1992). Expression of odorant receptors in spatially restricted subsets of chemosensory neurons. *NeuroReport*, *3*, 1053–1056.

Sullivan, S. L., Adamson, M. C., Ressler, K. J., Kozak, C. A., & Buck, L. B. (1996). The chromosomal distribution of mouse odorant receptor genes. *Proc. Natl. Acad. Sci. USA*, *93*, 884–888.

Tanabe, T., Iino, M., Ooshima, Y., & Takagi, S. F. (1974). Olfactory area in prefrontal lobe. *Brain Res.*, *80*, 127–130.

Thommesen, G., & Doving, K. B. (1977). Spatial-distribution of EOG in rat: Variation with odor quality. *Acta Physiol. Scand.*, *99*, 270–280.

Tsuboi, A., Yoshihara, S., Yamazaki, N., Kasai, H., Asai-Tsuboi, H., & Komatsu, M., et al. (1999). Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighboring glomeruli on the olfactory bulb. *J. Neurosci.*, *19*, 8409–8418.

Uchida, N., Takahashi, Y. K., Tanifuji, M., & Mori, K. (2000). Odor maps in the mammalian olfactory bulb: Domain organization and odorant structural features. *Nat. Neurosci.*, *3*, 1035–1043.

Ueki, S., & Domino, E. F. (1961). Some evidence for a mechanical receptor in olfactory function. *J. Neurophysiol.*, *24*, 12–25.

Vassar, R., Chao, S. K., Sitcheran, R., Nunez, J. M., Vosshall, L. B., & Axel, R. (1994). Topographic organization of sensory projections to the olfactory-bulb. *Cell*, *79*, 981–991.

Vassar, R., Ngai, J., & Axel, R. (1993). Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell*, *74*, 309–318.

Wellis, D., Scott, J., & Harrison, T. (1989). Discrimination among odorants by single neurons of the rat olfactory bulb. *J. Neurophysiol.*, *61*, 1161–1177.

Wesson, D. W., Carey, R. M., Verhagen, J. V., & Wachowiak, M. (2008). Rapid encoding and perception of novel odors in the rat. *PLoS Biol.*, *6*, e82.

Wilson, D. A. (1997). Binaral interactions in the rat piriform cortex. *J. Neurophysiol.*, *78*, 160–169.

Wilson, D. A. (2001). Receptive fields in the rat piriform cortex. *Chem. Senses*, *26*, 577–584.

Wilson, D. A. (2003). Rapid, experience-induced enhancement in odorant discrimination by anterior piriform cortex neurons. *J. Neurophysiol.*, *90*, 65–72.

Wilson, D. A., Best, A. R., & Sullivan, R. M. (2004). Plasticity in the olfactory system: Lessons for the neurobiology of memory. *Neuroscientist*, *10*, 513–524.

Wilson, M., & Bower, J. M. (1992). Cortical oscillations and temporal interactions in a computer-simulation of piriform cortex. *J. Neurophysiol.*, *67*, 981–995.

Wilson, D. A., & Stevenson, R. J. (2003). Olfactory perceptual learning: The critical role of memory in odor discrimination. *Neurosci. Biobehav. Rev.*, *27*, 307–328.

Wilson, D. A., & Stevenson, R. J. (2006). *Learning to smell.* Baltimore: Johns Hopkins University Press.

Wise, P. M., & Cain, W. S. (2000). Latency and accuracy of discriminations of odor quality between binary mixtures and their components. *Chem. Senses*, *25*, 247–265.

Wysocki, C. J., Pierce, J. D., & Gilbert, A. N. (1991). Geographic, cross-cultural, and individual variation in human olfaction. In T. V. Getchell (Ed.), *Smell and taste in health and disease* (pp. 287–314). New York: Raven Press.

Yokoi, M., Mori, K., & Nakanishi, S. (1995). Refinement of odor molecule tuning by dendrodendritic synaptic inhibition in the olfactory bulb. *Proc. Natl. Acad. Sci. USA*, *92*, 3371–3375.

Zald, D. H., & Pardo, J. V. (1997). Emotion, olfaction, and the human amygdala: Amygdala activation during aversive olfactory stimulation. *Proc. Natl. Acad. Sci. USA*, *94*, 4119–4124.

Zangrossi, H., Jr., & File, S. E. (1992). Behavioral consequences in animal tests of anxiety and exploration of exposure to cat odor. *Brain Res. Bull.*, *29*, 381–388.

Zelano, C., Bensafi, M., Porter, J., Mainland, J., Johnson, B., Bremner, E., Telles, C., Khan, R., & Sobel, N. (2005). Attentional modulation in human primary olfactory cortex. *Nat. Neurosci.*, *8*, 114–120.

Zhao, H. Q., Ivic, L., Otaki, J. M., Hashimoto, M., Mikoshiba, K., & Firestein, S. (1998). Functional expression of a mammalian odorant receptor. *Science*, *279*, 237–242.

23 Auditory Masking with Complex Stimuli

VIRGINIA M. RICHARDS AND GERALD KIDD, JR.

ABSTRACT The detection of a target sound embedded in a chaotic acoustical environment is a basic yet poorly understood component of auditory perception. This chapter reviews three aspects of such auditory masking, building toward the important problem of the perception of hearing a speech signal in the presence of competing speech sounds. First, the history of psychoacoustic masking experiments and the development of energy-based models to account for the resulting data are described. Then experiments and models of the detection of a tonal signal masked by randomly drawn maskers, an example of informational masking, are described. Informational masking experiments such as these are important because they reveal masking phenomena that are mediated more centrally than the masking associated with traditional masking studies. Finally, the roles of peripheral and central masking for the detection of speech masked by other speech sounds are discussed.

Imagine the following: You and a friend are standing on the corner of 34th and Chestnut Streets in Philadelphia, awaiting a bus to take you to Monks, a local pub known for Belgian-style beers. It is just after 5:00 P.M., and the traffic is heavy. And as it happens, a garbage truck is clearing public trash bins along Chestnut Street. Needless to say, you have to yell to be heard over the din of activity. This is an example of masking.[1] To communicate your message acoustically, you must broadcast your signal at a high intensity. Sound pressure waves superimpose, or add, in space, and what enters the ear is the accumulation of the sound in the environment. In enclosed environments—a classroom, for example—superposition includes not just the noise of students, audiovisual equipment, and so on, but also the sounds reflected off of walls, chalkboards, tables, and other surfaces. It is as though we live in an acoustical hall of mirrors. The auditory system has adapted to this din, allowing the segregation of a target sound away from the masking sounds. Efforts to understand this fundamental aspect of auditory perception form the basis of modern research in auditory masking.

In 1958, Tanner suggested that the intuitive sense of masking conveyed in the scenario described above is a vastly incomplete description of auditory masking. Tanner began with Licklider's (1951) description of masking: "Masking is thus the opposite of analysis; it represents the inability of the auditory mechanism to separate the tonal stimulation into components and discriminate between the presence or absence of one of them" (Tanner, 1958, p. 191; Licklider, 1951, p. 1005). This definition reflects the fact that the auditory system represents sounds *tonotopically*, such that the different frequencies of impinging sounds are encoded and represented by different populations of neurons. The failure of analysis noted by Licklider, then, essentially equates masking with limitations in the frequency resolution of the auditory system. When the signal and masker share common frequencies, the overlap of signal and masker energy makes the detection of the signal difficult because the shared energy is represented by the same population of neurons. Tanner noted that when the frequency of the tone to be detected is not fixed but chosen at random, the detectability of the signal decreases (Tanner & Norman, 1954). It is unlikely that this reduction in sensitivity reflects a failure of analysis at the auditory periphery. Should this decline in sensitivity be considered a form of masking? Tanner further pondered the detection of a tone of known frequency masked either by another tone or by Gaussian (white) noise. Should masking be defined independently of the properties of the masker? Regardless of whether the frequency of the tone to be detected is uncertain or the characteristics of the masker are varied, the end result is the same: A listener's ability to detect the signal changes. Should all of these examples be described by using the single term *masking*?

Tanner's point might be described in a slightly different way: As the sound stream is processed by the ascending auditory pathway, what information is lost and where in the processing is it lost? With regard to what masking tells us about the auditory system, the question becomes: How is it that biological systems effortlessly detect and follow an ongoing target sound under the pressure of multiple acoustical distracters? While not directly addressed in the current chapter, it is important to appreciate that for many individuals with compromised auditory capabilities, there is a decline in the ability to detect a target sound in the presence of other sounds especially when the listening situation is

VIRGINIA M. RICHARDS Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania
GERALD KIDD, JR. Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts

complex and uncertain. Presumably, when a more complete understanding of masking is achieved, advances in prosthetic devices such as hearing aids will follow.

In the current chapter, the answers to the complex question Tanner posed are not wholly addressed. Rather, the goal is to provide an update, asking: What are the peripheral and central constraints that yield auditory masking? Following the tradition in which Tanner worked, this chapter considers psychophysical data in which quantitative models associated with signal detection theory (cf. Green & Swets, 1974) are used to link the physical description of sounds with behavioral measurements. Thus, the scientific questions that are addressed are: (1) What are the peripheral constraints for the detection of a target among distracters? (2) What is known of the central constraints for the detection of a target among distracters? and (3) Given the large amount of masking associated with central processing, how does this affect our ability to recognize a target speech sound among competing speech? These questions are addressed in sequence, with an emphasis on the latter two.

Traditional masking studies: Energetic masking

Wegel and Lane (1924) provided empirical estimates of detection thresholds for a signal tone that is masked by another tone of different frequency. They described their results in terms of the *amount of masking*. The amount of masking is the increase in sound pressure level that is required to detect a target tone (the signal) in the presence of a second tone (the masker) compared to the detectabililty of the target tone in quiet. Figure 23.1 shows the results of a similar experiment by Egan and Hake (1950). Although the results of Egan and Hake (1950) and those of Wegel and Lane (1924) are quite similar, Egan and Hake's data are displayed because they are less influenced by factors such as auditory "beats" that distort the masking pattern. In this experiment, a narrowband noise masker was centered at 410 Hz, and the amount of masking, in dB, was plotted as a function of the signal frequency. The three functions are for three different masker levels: 40, 60, and 80 dB sound pressure level. Wegel and Lane's and Egan and Hake's results revealed several important features of auditory masking. First, they found that when the masker is distant in frequency from the signal, there is little masking. Consistent with the tonotopic organization of the auditory system, it is only when the masker is near in frequency to the signal that masking occurs. Second, when the masker level is increased, the signal level too must be increased to maintain equal detectability. Third, when the masker has a frequency lower than the signal frequency, there is more masking than occurs for the converse

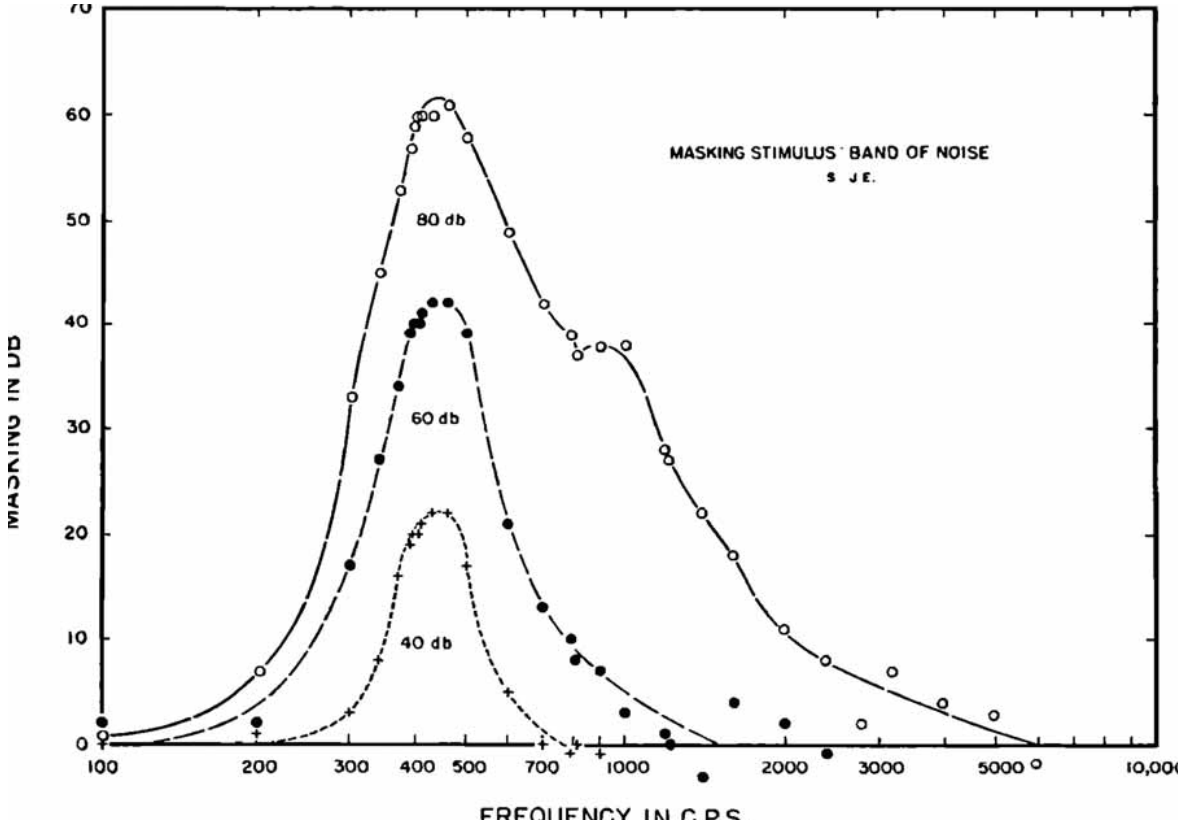


FIGURE 23.1 The amount of masking for a tonal signal masked by a narrowband noise centered at 410 Hz is plotted as a function of signal frequency. The level of the narrowband noise was 40, 60, and 80 dB sound pressure level. (From Egan & Hake, 1950.)

situation; for example, to detect a midfrequency signal, a low-frequency masker produces more interference than a high-frequency masker does. Fourth, they found that as the frequency of the masker increased, the frequency range over which the masker interfered with signal detectability increased. This work and many other related studies have provided psychoacoustical evidence corresponding to the frequency selectivity that is found at the level of the basilar membrane within the cochlea as well as suggesting the potential that the transduction chain includes nonlinearities.

Fletcher (1940) provided early estimates of the effective bandwidth of interaction at the auditory periphery. In one of his experiments, the task was to detect a tone added to a noise masker. The bandwidth of the masker, which was centered at the signal frequency, was systematically broadened. The idea was that as the noise bandwidth increased, the detectability of the added tone should decrease. This is because the masker power passed through the *internal* bandpass filter[2] (currently referred to as the *auditory filter*) would increase with increasing masker bandwidth, causing more and more masking of the signal tone. Once the bandwidth exceeded the bandwidth of the internal filter, no further increase in masking would be expected. As such, when the bandwidth of the masker increased, Fletcher expected the detectability of the added tone to first decrease and then level off. This prediction held true, allowing Fletcher to estimate the bandwidth of the effective internal filter.

Subsequent work has refined the assumptions upon which Fletcher's model was based, and this contemporary model is referred to as the *power spectrum model of masking*. By this model, for the detection of a tone added to any masker, noise or otherwise, a listener monitors the output of the single auditory filter. Work by Patterson and colleagues and Moore and colleagues (e.g., Patterson, 1976; Patterson, Nimmo-Smith, Weber, & Milroy, 1982; Glasberg & Moore, 1990; Moore, 1995; Unoki, Irino, Glasberg, Moore, & Patterson, 2006) has provided computational/psychophysical approaches that allow for the estimation not only of the bandwidth but also the shape of auditory filters. The assumptions associated with the power spectrum model of masking include the following: (1) Listeners monitor the output of a single auditory filter, the filter with the highest signal-to-noise ratio; (2) the auditory system can be modeled as a set of contiguous overlapping bandpass (and notably, nonlinear) filters[2]; (3) listeners detect the added tone at a constant signal-to-noise ratio; and (4) there is variance in the listener's auditory/decision-making system. The last assumption groups together all possible sources of error, including neural jitter, uncertainty regarding the true characteristics of the signal to be detected, variations in motivation, and so on. The resulting estimates of auditory filter characteristics provide a convenient psychophysical tool for studying peripheral interactions. The masking associated with limitations in analysis at the auditory periphery is often referred to as *energetic masking*, a term coined by Pollack (1975) as a contrast to *informational masking*.

Informational masking: Simultaneous multitone maskers

Different from energetic masking, central masking, or *informational masking*, occurs when there is an elevation of threshold due to stimulus uncertainty (cf. Leek, Brown, & Dorman, 1991; Neff, 1995). An alternative description is that informational masking is the amount of masking above and beyond energetic masking or the masking that occurs even when there is insufficient interaction between the signal and masker at the auditory periphery to explain the results (see Durlach et al., 2003a, for a discussion).

One means of inducing informational masking is to use maskers composed of several tones, with the frequencies of the tones chosen at random on each presentation. This procedure was introduced by Neff and Green (1987). The signal to be detected was fixed in frequency (e.g., 1000 Hz), and the number of tones that composed the masker ranged, in different blocks, from 1 to 100. Critically, the frequencies of the masker tones were drawn at random prior to each stimulus presentation, creating masker uncertainty. To understand the difficulty that this task poses, imagine trying to hear a 1000-Hz tone embedded in a masker, but the masker's timbre varies vastly from presentation to presentation. As a simple example, the masker might shift randomly from a piano to an oboe to an organ from trial to trial. There is little basis for comparison among the sounds, making it difficult to detect the added tone. This is the core of informational masking: Under pressures of masker variation, listeners fail to *selectively attend* to the known signal frequency.

An empirical example relating the amount of informational masking to the number of masker tones composing the masker is shown in figure 23.2 (Neff & Green, 1987). Two aspects of this graph deserve note. First, even when the masker was a single sinusoid (one masker component), thresholds were on average 20 dB higher than those when no masker was present. Second, thresholds were highest when the masker was composed of 10–30 or so components. With regard to understanding this function, it should be noted that energetic masking is expected to increase as the number of tones increases. This reflects the statistics of the masker. The more masking tones there are, the more likely it is that one or more of the masker tones' frequency will be close to the signal frequency. With regard to the nonmonoticity of the function, Oh and Lutfi (1998) have developed a model that captures this general shape. As a brief description, their model suggests that the more variance there is at the output of each auditory filter, computed across trials, the larger is the amount of informational masking. When there are few masker components or many masker components, the variance in any one auditory filter is modest.

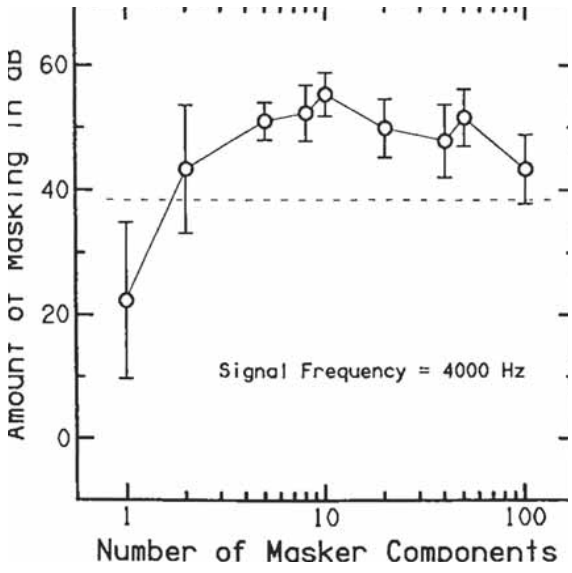


FIGURE 23.2 Amount of masking is plotted as a function of the number of tones comprising the masker. The dashed line is the amount of masking for Gaussian noise. (From Neff & Green, 1987.)

An additional common finding in informational masking experiments is large individual differences in levels of performance. In many instances, the range in the amount of masking approaches 50–60 dB (Neff & Dethlefs, 1995). This range represents approximately one-half of the dynamic range of the auditory system, which, depending on the way in which it is computed, typically is greater than 100 dB. Currently, the reasons for this enormous range of performance across individual subjects are poorly understood.

Even when individual differences are noted, it is clear that random multitone maskers are tremendously effective maskers of tonal signals. Additionally, it is apparent that informational masking reflects limitations in central processing. As a result of the large amount of masking, modest changes in cognitive processing can be reliably probed. For this reason, researchers have come to ask, in essence, how informational masking can be reduced, or *released*. To answer this question, it is critical that tools exist to separate peripheral from central processing (cf. Brungart, Chang, Simpson, & Wang, 2006).

With regard to potential means of releasing informational masking, an initial step is to recognize that informational masking must surely reflect, to a significant degree, the auditory system's grouping together of the signal and masker. That is, the signal and masker are, to a large degree, perceived as emanating from a single sound source. The factors that are associated with this grouping mechanism contribute to the failure of the auditory system to "hear out," or potentially selectively attend to, the signal tone. Consistent with this result is that listeners with superior sensitivity rely most heavily on information in the region of the signal, whereas listeners with poorer sensitivity tend to integrate information from frequency regions that are distant from the signal frequency (e.g., Alexander & Lutfi, 2004; Oh & Lutfi, 1998; Richards, Tang, & Kidd, 2002; Richards & Neff, 2004; Tang & Richards, 2003). Drawing a rough parallel with Treisman and colleagues' feature integration theory (e.g., Triesman & Gelade, 1980), one potential description is as follows. When the masker is composed of relatively few masker tones (say, 10), the masker tones and the signal are separately analyzed at the auditory periphery, yielding several tones with common characteristics or features. Because the auditory system cannot distinguish an increase from 10 tones (masker alone) to 11 tones (masker plus signal), the signal is very difficult to detect. By this notion, introducing differences in the characteristics of the signal compared to the masker components ought to allow the signal to be segregated from the masker, thereby releasing masking. In fact, this result has been obtained (e.g., Neff, 1995; Kidd, Mason, & Arbogast, 2002; Durlach et al., 2003b). For example, if the signal to be detected starts shortly after the masker or is a frequency glide rather than a simple tone, then informational masking is substantially reduced. At present, the relationship between the amounts of release from informational masking due to changes in the characteristics of the signal versus the masker (and masker components) is not well understood.

The amount of informational masking can also be reduced by providing listeners with a cue that reduces stimulus uncertainty. Several types of cues have been tested: signal cues, masker cues, and signal-plus-masker cues (e.g., Richards & Neff, 2004; Richards, Huang, & Kidd, 2004). These cues are exact replicates of the stimuli tested on each trial. A signal cue reminds the listener of the signal's frequency thereby reducing signal-frequency uncertainty. The available data indicate that when signal cues are tested, listeners tend to rely on or heavily weight the region near the signal frequency (Richards & Neff, 2004). Similarly, masker cues would be expected to reduce masker frequency uncertainty, while signal-plus-masker cues would be expected to reduce uncertainty for both. Note that masker and signal-plus-masker cues reduce stimulus uncertainty nearly equally, so they might be expected to provide approximately equal amounts of release from masking. The data, however, which are discussed below, indicate otherwise.

Richards and colleagues (Richards & Neff, 2004; Richards et al., 2004) have completed several experiments examining the efficacy of different cue types in their ability to release informational masking. Richards and Neff examined the impact of providing a cue, both signal and masker cues, before the detection trial. The signal, when present, was either a 1000-Hz tone or a randomly drawn tone for each signal trial, depending on the condition that was tested.

The masker was composed of six tones whose frequencies were randomly drawn. Masked thresholds (signal level required for a d' of approximately 1) were estimated both with and without pretrial cues, and data collection was blocked to allow listeners to take full advantage of the cue. Listeners were highly trained and were provided with feedback following each trial.

First consider the results with a signal cue. Regardless of whether the signal frequency was fixed or random, thresholds were lower (superior sensitivity) when a cue preceded a trial than when there was no cue. Moreover, cued thresholds were only slightly poorer when the signal frequency was random than when it was fixed. Surprisingly, there was no correlation between the informational masking threshold and the amount of release from informational masking. That is, although there were substantial individual differences in the amount of informational masking, the release from informational masking provided by the signal cue was approximately equal across listeners.

Next consider the results with a masker cue. A preview of the masker provides information as to which frequency regions (masker tones) *not* to pay attention to on the subsequent trial. The results indicated that a masker cue was as effective as the signal cue in releasing informational masking. In one direct comparison, both signal and masker cues provided 10–15 dB of release from informational masking. Additional data showed that the benefit of both signal and masker cues lasts a substantial length of time. For example, delays between the cue and the trial ranging from 50 to 500 ms were equally effective. To summarize, the results of Richards and Neff (2004) indicated that a preview of either the signal or the masker can provide a substantial release from informational masking. However, neither a signal nor a masker cue fully released informational masking. It was estimated that on average, approximately 30 dB of informational masking still remained.

The fact that a preview of the masker allows a release from informational masking is consistent with the phenomenon of auditory enhancement, which in turn is consistent with the general finding that the auditory system is adept at detecting when something "new" is added to the sound stream (i.e., the introduction of new sound sources). Auditory enhancement is a term that describes the fact that the detection of a signal can be *enhanced* when energy "flanking" the signal frequency precedes the signal (e.g., Viemeister, 1980; Viemeister & Bacon, 1982; Wright, McFadden, & Champlain, 1993). Whether cueing effects in informational masking experiments reflect the same mechanisms as for the auditory enhancement reported in these studies remains to be determined. However, the results that are considered next suggest that they do contribute.

The importance to the auditory system of "new" information can be evaluated by measuring the release from informational masking for *pretrial* cues versus *posttrial* cues. The logic is as follows. If a masker cue is effective in releasing informational masking, regardless of whether it precedes or follows a trial, it would suggest that the release from information masking reflects a reduction in uncertainty. In contrast, if the pretrial cue is much more effective than the posttrial cue then, for the masker-only cue at least, there would be support for the hypothesis of auditory enhancement. By the same token, a signal-plus-masker cue should also reduce uncertainty and thus should release informational masking but would not produce auditory enhancement. Figure 23.3 shows the results of two informational masking experiments that examine these predictions (Richards et al., 2004). First consider the open symbols to the left. The open square shows the average value of d' for a signal of fixed frequency and fixed level when a masker cue preceded each trial. The open circle (hidden by the solid symbols) shows d' for the same signal but with a preview of the signal-plus-masker stimulus. The masker cue provides substantially more release from masking than the signal-plus-masker cue does, consistent with the idea that the detection of something new (i.e., when the following trial has a signal) is enhanced by the auditory system.

Second, consider the open symbols to the right in figure 23.3. In this case, the cue followed the detection trial. As such, it is the signal-plus-masker cue that provides something "new" to be detected (i.e., a masker trial followed by a signal-plus-masker cue). The values of d' for the signal-plus-masker and masker posttrial cues are approximately the same. Moreover, the values of d' are approximately intermediate between the values that were obtained for the pretrial cues.

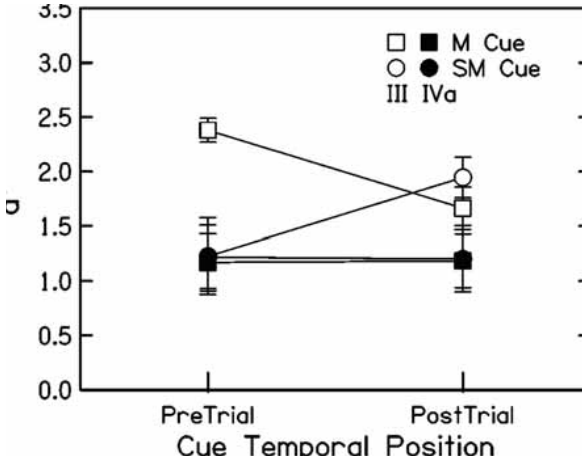


FIGURE 23.3 The index of detectability, d', is plotted for two informational masking experiments. For one experiment, the signal frequency was fixed (open symbols), and for the other, the signal frequency was random (solid symbols). The cue was either a copy of the masker (square) or signal-plus-masker (circle), and the cue was presented either before or after the trial (abscissa). (From Richards et al., 2002.)

Across several experiments, data such as these indicate that a preview of the masker provides substantial release from informational masking, while the three remaining cue conditions (pretrial signal-plus-masker, posttrial masker, and signal-plus-masker) provide approximately equal, but less, release from informational masking. This result suggests that the larger release from informational masking does not simply reflect the temporal introduction of a "new" sound; pretrial masker cues and posttrial signal-plus-masker cues introduce an equal number of "new" sounds to be detected. This result contrasts with visual search studies in which pretrial and posttrial cues can be equally effective (cf. Kinchla, Chen, & Evert, 1995), potentially a situation in which the cue effectively reduces uncertainty.

Next consider the solid symbols in figure 22.3. In this case, the signal frequency was random but of the same level as for the conditions described above. When the signal frequency is random, all four types of cues (masker and signal-plus-masker by pretrial and posttrial positions) are equally effective.

Clearly, cues do provide a release from informational masking, but the pattern of results is complex. The dominant advantage of a masker cue is restricted to situations in which the signal frequency is *known* and the cue precedes the detection trial, at least for the temporal parameters tested by Richards and colleagues. These results are not wholly consistent with the hypothesis that the release from informational masking associated with a masker pretrial cue reflects the auditory systems ability to detect a new sound source. Nor are they consistent with the notion that masker and signal-plus-masker cues simply reduce uncertainty.

These results and others indicate the complexity of forms of masking that occur even for relatively simple psychoacoustic tasks. Indeed, informational masking is likely to be very prevalent in many psychoacoustic tasks. For example, Lutfi (1990) estimated that even for the traditional task of detecting a tone added to Gaussian noise, some 20% of the total masking is informational masking. Next, we discuss the ways in which research concerning informational masking contributes to an understanding of masking in a far more complex environment: the recognition of target speech sounds in the presence of competing utterances.

Informational masking and speech recognition

The acoustic signals that we use to communicate with most frequently are speech sounds. Other sounds may interfere with our ability to understand speech, and this interference may also be considered in terms of energetic and informational masking. Early studies of the masking of speech—similar to the early studies of the masking of tones mentioned above—often used Gaussian noise or filtered Gaussian noise as a masker. The interpretation of the results of speech recognition studies in the presence of noise was straightforward: The noise obscured or covered up portions of the speech sounds, and this loss of information had effects on intelligibility that were quite predictable. High-frequency noise, for example, might mask the high-frequency parts of a word such as the fricative consonant /s/. In such a case, the errors that would occur were likely to be errors involving the masked speech sound. In fact, the errors that occur from energetic masking can be accounted for by articulation index theory (French & Steinberg, 1947), which predicts how intelligible a speech signal will be on the basis of the signal-to-noise ratio in a number of frequency bands (cf. Allen, 2005). However, in contrast to the rather straightforward effects of noise (and some other types of energetic maskers) on speech intelligibility, sounds that produce informational masking—especially other talkers that are similar to the talker one wants to listen to—produce interference that might be difficult to predict and may have effects that are less straightforward than simple noise masking. The speech signal itself is complex and time-varying with frequency components that often change from moment to moment. When there are multiple sources of speech occurring simultaneously, the extent to which they overlap acoustically—and subsequently in the neural representations in the auditory system—also may vary from moment to moment. This time-varying overlap of the frequency spectra of multiple speech signals means that quantifying the amount of energetic masking that is present can be challenging, especially in natural listening situations.

It has long been appreciated that the effect of one speech signal on the intelligibility of another speech signal can go well beyond energetic masking. Broadbent (1952) provided an early and clever demonstration of this. In his study, a target sentence was broken up into individual words, and irrelevant words from another talker were interspersed in between the target words. Thus the words from the two talkers alternated in time. No energetic masking occurred because the target and masker words did not overlap. However, significant interference was observed in subjects' ability to report the target speech. Broadbent considered this result to reflect a failure of selective attention. However, not all possible intervening sounds produce this type of informational masking. Recently, Kidd, Best, and Mason (2008) demonstrated that playing bursts of noise or time-reversed masker speech in between the target speech has only minor effects on intelligibility. Thus the meaningfulness of the masker speech and its similarity to the target speech are crucial for producing informational masking.

Carhart, Tillman, and Greetis (1969) concluded that some types of maskers—specifically, modulated noise or speech—produced masking of target speech that went beyond energetic masking. They found that combining these complex maskers produced masking that could not be reconciled with

the individual effects of each masker separately and termed the extra masking *perceptual masking*. To account for this extra masking, they speculated that more than one process was necessary and noted other examples involving the masking of simpler stimuli in which "excess additivity" was observed. Although the terms *energetic* and *informational masking* were not introduced until later (Pollack, 1975), these speech masking results—similar to the multitone masking results above—seemed to require the actions of separate mechanisms, one peripheral and the other central. Furthermore, the central factors contributing to masking likely involve multiple processes. Even though each may be considered a type of informational masking, there is a need to understand and distinguish among these processes (cf. Kidd, Mason, Richards, Gallun, & Durlach, 2008).

As was described above, one means for demonstrating masking effects in speech recognition in the absence of any peripheral overlap of target and masker was devised by Broadbent (1952) using the alternating-word paradigm. A more recent experiment used signal processing to achieve similar, if not as complete, isolation of informational masking in speech identification. This study by Arbogast and colleagues (2002) will be discussed in some detail because it illustrates clearly the distinction between energetic and informational masking of speech, and the stimuli resulting from the signal processing they employed provide some interesting parallels to the multitone masking stimuli discussed above.

Figure 23.4 illustrates the steps in processing target and masker speech. The speech materials used were developed by Bolia, Nelson, Ericson, and Simpson (2000) and are referred to as the coordinate response measure (CRM). The CRM is a closed-set speech identification test in which the observer must report certain key words that are colors and numbers. As shown in figure 23.4, these recorded sentences were first given a gradual high-frequency emphasis, then filtered into 1/3-octave bands, half-wave rectified and low-pass filtered. This process extracts the amplitude envelope within each band. These envelope functions are then used to modulate pure-tone carriers corresponding to the center frequencies of each of the bands. Only a few such envelope-modulated carrier tones need to be combined to form intelligible speech. Thus a target sentence and a masker sentence may be composed of mutually exclusive frequency bands, but each retains a high degree of intelligibility.

Figure 23.5 shows two such sets of bands for two different sentences. These magnitude spectra, which are averaged over the length of each sentence, are characterized by sets of very narrow frequency bands that overlap only many decibels below the peaks. The effect of this processing is to render highly intelligible target and masker sentences into sets of narrow frequency bands that interact minimally with respect to energetic masking. This latter assertion was supported by a control condition in which one of the sentences was replaced by sets of matched narrow bands of noise (not shown in figure 23.5; see Arbogast et al., 2002). The difference in the amount of masking produced by processed speech and by the noise control was taken to be an estimate of the amount of informational masking produced by the speech. Figure 23.6 illustrates this result.

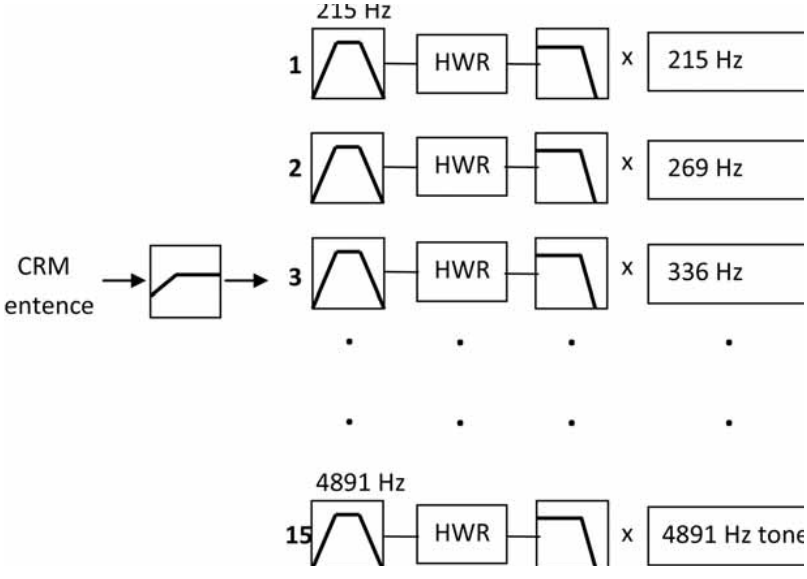


Figure 23.4 A schematic illustration of the steps in processing the target and/or masker speech into sets of narrow bands. The various stages include (left to right): (1) gradual high-frequency emphasis, (2) filtering the speech into one-third-octave bands, (3) half-wave rectifying the filtered speech, (4) low-pass filtering of the rectified waveforms to extract the envelopes, and (5) multiplying the envelope functions with pure-tone carriers centered in each frequency band. (Adapted from Arbogast, 2003.)

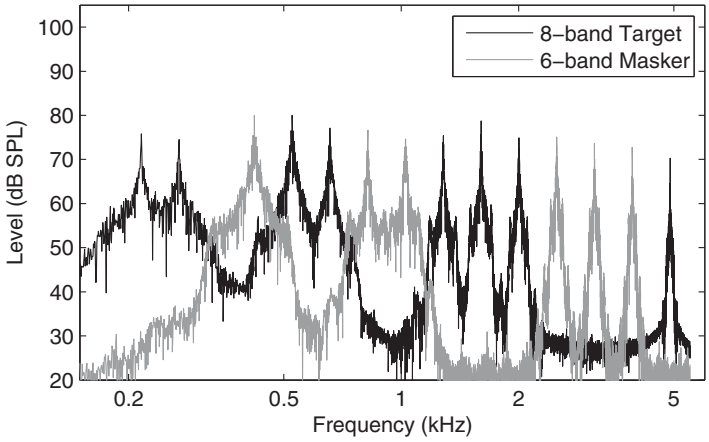


FIGURE 23.5 Magnitude spectra of target (black) and masker (gray) speech processed as shown in figure 23.4 into mutually exclusive frequency bands.

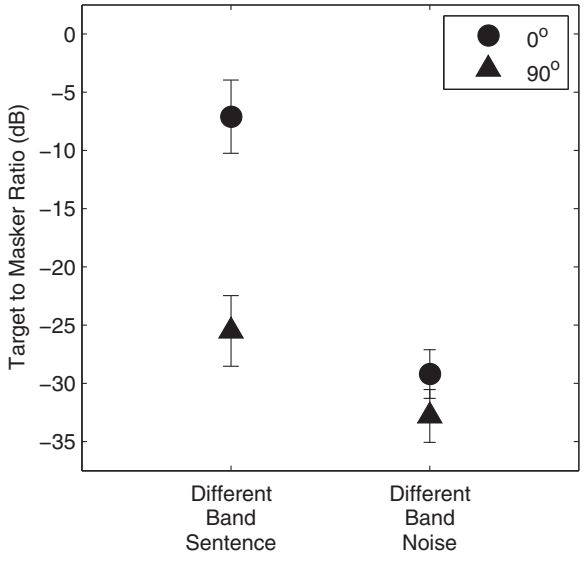


FIGURE 23.6 These are a portion of the results from Arbogast and colleagues (2002) replotted to emphasize the difference between energetic (different band noise) and informational (different band sentence) maskers. The values plotted are the level of the target relative to the level of the masker at speech reception threshold (50% correct identification).

The data plotted in this figure are group mean speech reception thresholds that indicate the level of the target relative to the level of the masker that allowed the listeners to correctly identify the key words in the target 50% of the time (chance performance is 3% correct). The two values on the left side of the graph represent the thresholds for the speech masker (called *different band sentence* because the frequency bands of the masker sentence are different from those of the target sentence), while the two values on the right side of the graph represent the thresholds for the noise masker control (called *different band noise*). For each masker, the higher of the two points indicates a threshold that is obtained when the target and masker were both played from the same loudspeaker, which was located directly in front of the listener (0° spatial separation). The lower of the two points in each case indicates a threshold that is obtained when the target was presented from a loudspeaker directly in front of the listener while the masker was presented from a second loudspeaker located directly to the right of the listener (referred to as a spatial separation of 90°).

There are two main points to be made regarding the design of this experiment and the results shown in figure 23.6. First, the goal in the processing of speech in this case is very similar to the goal in the design of the multitone masking experiment discussed above. The masker energy is positioned in frequency regions that are remote from the target energy to reduce the peripheral overlap of excitation of the two minimizing energetic masking. The assumption is that a large proportion of the masking that occurs is informational masking that is not due to peripheral overlap of excitation. For both types of maskers—multitone and multiband speech—control conditions have been examined that support this interpretation (for the multitone masker, see Durlach et al., 2005). In the present case, the different band noise is the energetic masking control for the different band sentence masker and the difference in masking each produces is an estimate of the amount of informational masking. In this case, there are two comparisons: one when the target and masker are colocated and the second when they are spatially separated. From figure 23.6, the former comparison (circles) yields about 22 dB of informational masking, while the latter (triangles) yields about 7 dB. Second, when large amounts of informational masking are produced, stimulus manipulations that cause or strengthen the perceptual segregation of the target from the masker reduce the observed masking. In figure 23.6, it is obvious that spatially separating the speech target and speech masker produced a large decrease in the amount of masking (approximately 18.4 dB)

primarily through a reduction in informational masking. Although some energetic masking release could be produced by spatial separation of sources, the control condition of multiband noise (right side of figure) indicates that this was a minor factor in the experiment. This prominent role of perceptual segregation in reducing informational masking, but not energetic masking, has been reported in several other studies (cf. Neff, 1995; Kidd, Mason, Rohtla, & Deliwala, 1998; Freyman, Helfer, McCall, & Clifton, 1999).

Concluding comments

The study of masking over the past century has provided important information regarding the functional organization of the auditory system. Consistent with this history, in recent years, studies of informational masking have provided a potent tool for an understanding of the relative contributions of peripheral and central processing in auditory perception. The ultimate goal is clear: to more fully understand *why* we hear *what* we hear in the cluttered acoustical environment. While the psychoacoustics research summarized above is incomplete—there are many fundamental questions to be answered, including a comprehensive answer to that posed by Tanner (1958)—studies of masking continue to offer a framework in which questions regarding the cognition of hearing can be addressed.

ACKNOWLEDGMENTS We acknowledge the assistance of Rong Huang and Christine Mason in preparing this chapter. This work was supported by grants RO1 DC02012 and RO1 DC04545 from the National Institutes of Health and grant FA9950-50-1-2005 from the Air Force Office of Scientific Research.

NOTES

1. It might happen that you are wearing a new rain hat with a wide brim. This hat leads to reflections of the sounds as they enter your ear, reflections that you have not experienced before. Alternatively, an impending head cold might clog your right Eustachian tube, providing an imbalance in the sounds transmitted by your right and left ears. These challenges, which point to the efficiency with which the auditory system recalibrates, are beyond the scope of the current chapter.
2. A bandpass filter is a device that allows a range of frequencies to pass but attenuates, or rejects, other frequencies. As an illustrative concept, a bandpass filter centered at 1000 Hz with a passband (or bandwidth) of 300 Hz would allow all frequencies between 850 and 1150 Hz to pass unimpeded yet would attenuate frequencies lower than 850 Hz or higher than 1150 Hz. For a hypothetical rectangular filter, the frequencies between 850 and 1150 Hz would pass with no change in level or phase. Moreover, the output of a linear filter for two tones would be equal to the sum of the output of the filter for each tone alone.

   Realistic biological, analog, and digital filters do not match this ideal. Typically, bandpass filters are not rectangular. For the example provided above, a rectangular filter would completely attenuate a 849-Hz tone but would pass a 851-Hz tone without attenuation. Instead, filters have "skirts" such that there is a gradual change in attenuation as a function of frequency from unattenuated to fully attenuated. Realistic filters also alter the phase of tones that pass through the filter, a phase shift that depends on the frequency of the tone. Finally, nonlinear filters, which are used to describe the auditory periphery, might not have unity gain. Depending on the intensity of the incoming sound, the filter might amplify the sound.

   There are many examples of filters in everyday life. The telephone, for example, does not pass all frequencies; low and high frequencies are vastly attenuated. Our auditory system provides another example: Try as one might, the system does not pass/represent very high frequencies (e.g., above 20 kHz or so).

REFERENCES

ALEXANDER, J. M., & LUTFI, R. A. (2004). Information masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *J. Acoust. Soc. Am.*, *116*, 2234–2247.

ALLEN, J. B. (2005). Consonant recognition and the articulation index. *J. Acoust. Soc. Am.*, *117*, 2212–2223.

ARBOGAST, T. L. (2003). *The effect of spatial separation on informational and energetic masking of speech in normal-hearing and hearing-impaired listeners*. Doctoral dissertation, Boston University.

ARBOGAST, T. L., MASON, C. R., & KIDD, G., Jr. (2002). The effect of spatial separation on informational and energetic masking of speech. *J. Acoust. Soc. Am.*, *112*, 2086–2098.

BOLIA, R. S., NELSON, W. T., ERICSON, M. A., & SIMPSON, B. D. (2000). A speech corpus for multitalker communications research. *J. Acoust. Soc. Am.*, *107*, 1065–1066.

BROADBENT, D. E. (1952). Failures of attention in selective listening. *J. Exp. Psychol.*, *44*, 428–433.

BRUNGART, D. S., CHANG, P. S., SIMPSON, B. D., & WANG, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *J. Acoust. Soc. Am.*, *120*, 4007–4018.

CARHART, R., TILLMAN, T. W., & GREETIS, E. (1969). Perceptual masking in multiple sound backgrounds. *J. Acoust. Soc. Am.*, *45*, 694–703.

DURLACH, N. I., MASON, C. R., GALLUN, F. J., SHINN-CUNNINGHAM, B., COLBURN, H. S., & KIDD, G., Jr. (2005). Inforrmational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. *J. Acoust. Soc. Am.*, *118*, 2482–2497.

DURLACH, N. I., MASON, C. R., KIDD, G., Jr., ABORGAST, T. L., COLBURN, H. S., & SHINN-CUNNINGHAM, B. G. (2003a). Note on informational masking. *J. Acoust. Soc. Am.*, *113*, 2984–2987.

DURLACH, N. I., MASON, C. R., SHINN-CUNNINGHAM, B. G., ARBOGAST, T. L., COLBURN, H. S., & KIDD, G., Jr. (2003b). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J. Acoust. Soc. Am.*, *114*, 368–379.

EGAN, J. P., & HAKE, W. H. (1950). On the masking pattern of a simple auditory stimulus. *J. Acoust. Soc. Am.*, *22*, 622–630.

FLETCHER, H. (1940). Auditory patterns. *Rev. Mod. Phys.*, *12*, 47–65.

FRENCH, N. R., & STEINBERG, J. C. (1947). Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.*, *19*, 90–119.

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am.*, *106*, 3578–3588.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.*, *47*, 103–138.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics.* Huntington, NY: Robert E. Krieger.

Kidd, G., Jr., Best, V., & Mason, C. R. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech. *J. Acoust. Soc. Am*, *124*, 3793–3802.

Kidd, G., Jr., Mason, C. R., & Arbogast, T. L. (2002). Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *J. Acoust. Soc. Am.*, *111*, 1367–1376.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 143–190). New York: Springer Science+Business Media.

Kidd, G., Jr., Mason, C. R., Rohtla, T. L., & Deliwala, P. S. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *J. Acoust. Soc. Am.*, *104*, 422–431.

Kinchla, R. A., Chen, Z., & Evert, D. (1995). Precue effects in visual search: Data or resource limited? *Percept. Psychophys.*, *57*, 441–450.

Leek, M. R., Brown, M. E., & Dorman, M. F. (1991). Informational masking and auditory attention. *Percept. Psychophys.*, *50*, 205–214.

Licklider, J. C. R. (1951). Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (1st edition, pp. 985–1039). New York: Wiley.

Licklider, J. C. R., (1963). Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology cell nature* (5th ed., pp. 985–1039). New York: Wiley.

Lutfi, R. A. (1990). How much masking is informational masking? *J. Acoust. Soc. Am.*, *88*, 2607–2510.

Moore, B. C. J. (1995). Frequency analysis and masking. In *Handbook of perception and cognition: Vol. 6. Hearing* (2nd ed., pp. 161–204). E. C. Carterette & M. P. Friedman (Series Eds.) & B. C. J. Moore (Vol. Ed.). San Diego: Academic Press.

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous random-frequency maskers. *J. Acoust. Soc. Am.*, *98*, 1909–1920.

Neff, D. L., & Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *J. Acoust. Soc. Am.*, *98*, 125–134.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Percept. Psychophys.*, *41*, 409–415.

Oh, E., & Lutfi, R. A. (1998). Nonmonotonicity of informational masking. *J. Acoust. Soc. Am.*, *104*, 3489–3499.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *J. Acoust. Soc. Am.*, *59*, 640–654.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram and speech threshold. *J. Acoust. Soc. Am.*, *72*, 1788–1803.

Pollack, I. (1975). Auditory informational masking. *J. Acoust. Soc. Am.*, *57*, S5.

Richards, V. M., Huang, R., & Kidd G., Jr. (2004). Masker-first advantage for cues in informational masking. *J. Acoust. Soc. Am.*, *116*, 2278–2288.

Richards, V. M., & Neff, D. L. (2004). Cuing effects for informational masking. *J. Acoust. Soc. Am.*, *115*, 289–300.

Richards, V. M., Tang, Z., & Kidd G., Jr. (2002). Informational masking with small set sizes. *J. Acoust. Soc. Am.*, *111*, 1359–1366.

Tang, Z., & Richards, V. M. (2003). Estimation of a linear model in an informational masking study. *J. Acoust. Soc. Am.*, *114*, 361–367.

Tanner, W. P. Jr. (1958). What is masking? *J. Acoust. Soc. Am.*, *30*, 919–921.

Tanner, W. P., Jr., & Norman, R. Z. (1954). The human use of information: II. Signal detection for the case of unknown signal parameters. *J. Trans. Inst. Radio Engrs. PGIT*, *4*, 222–227.

Triesman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cogn. Psych.*, *12*, 97–136.

Unoki, M., Irino, T., Glasberg, G., Moore, B. C. J., & Patterson, R. D. (2006). Comparison of the roex and gammachirp filters as representations of the auditory filter. *J. Acoust. Soc. Am.*, *120*, 1474–1492.

Viemeister, N. F. (1980). Adaptation of masking. In G. Van den Brink & F. A. Bilsen (Eds.), *Psychological, physiological and behavioral studies of hearing.* Delft, Netherlands: Delft University Press.

Viemeister, N. F., & Bacon, S. P. (1982). Forward masking by enhanced components in harmonic complexes. *J. Acoust. Soc. Am.*, *71*, 1502–1507.

Wegel, R. L., & Lane, C. E. (1924). The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.*, *23*, 266–285.

Wright, B. A., McFadden, D., & Champlin, C. (1993). Adaptation of suppression as an explanation of enhancement effects. *J. Acoust. Soc. Am.*, *94*, 72–82.

24 Insights into Human Auditory Processing Gained from Perceptual Learning

BEVERLY A. WRIGHT AND YUXUAN ZHANG

ABSTRACT Many auditory skills improve with practice, indicating malleability of the underlying neural system. Here we consider the effect of training on the human perception of basic sound attributes such as frequency, intensity, and duration. We first compare learning patterns across multiple tasks, in each of which listeners discriminate changes in a different sound attribute. These patterns differ markedly for different tasks and sometimes even for different stimuli within the same task, in terms of both how performance changes over training sessions and how learning generalizes to untrained conditions. The differences suggest that training on different tasks affects different neural processes. We then describe in more detail sets of training experiments on auditory-timing and spatial-hearing skills and make inferences about the underlying neural processes affected by the training. Finally, we speculate about the neural underpinnings of auditory learning. This chapter thus illustrates that the examination of auditory learning can provide unique insights into the human perception and neural processing of sounds.

A remarkable and often unrecognized characteristic of human perceptual abilities is that they can be improved with practice. Such perceptual learning indicates that the underlying neural processes are malleable. Investigations of the circumstances that yield this learning and of the patterns with which this learning occurs have both theoretical and practical value. On the theoretical side, this information provides insight into the architecture and plasticity of the neural processes that govern perceptual performance. On the practical side, it can guide the development of more effective and efficient perceptual training regimens to aid individuals with perceptual disorders as well as others who desire enhanced perceptual skills. To date, perceptual learning has been examined primarily in the visual system. Here, we instead describe select aspects of perceptual learning in the auditory system.

BEVERLY A. WRIGHT and YUXUAN ZHANG Department of Communication Sciences and Disorders and Interdepartmental Neuroscience Program, Northwestern University, Evanston, Illinois

To help establish the principles of auditory learning, we and others have focused our investigations on basic auditory skills and simple training regimens. In such cases, during training, listeners are asked to discriminate between small variations in only one attribute of a relatively simple sound, such as to determine which of two tones has a higher frequency or a longer duration. Establishing the learning patterns under these circumstances forms a baseline for interpreting improvements on more complex tasks and with more complex training regimens. Suggesting that this is a reasonable approach, we have recently seen that phenomena we first observed in learning on simple auditory tasks occur on speech-perception tasks as well. Similar reasoning has also guided a large number of investigations in the visual system (Rust & Movshon, 2005).

In this chapter, we begin with a brief review of differences across different trained tasks in the pattern of performance improvement resulting from training and argue that these differences constitute evidence for the involvement of different neural processes in these different cases. We then consider in more detail the learning patterns on auditory temporal and spatial tasks to illustrate how these patterns can be used to make inferences about the particular neural processes that were modified through training. We conclude with a brief discussion of the neural bases of auditory learning itself.

Evidence that different neural processes contribute to learning on different auditory tasks

The behavioral evidence that auditory learning involves different neural processes on different tasks arises primarily from examination of three aspects of the learning patterns: learning on the trained condition, across-task generalization, and across-stimulus generalization. This evidence rests on the basic assumption that the pattern of learning and generalization is determined by the particular neural circuitry that is being modified as well as by the particular type of modification that is occurring (e.g., Hochstein & Ahissar,

2002; Karni & Sagi, 1991). Given this assumption, differences in learning and generalization patterns across different tasks suggest that training on different tasks induced the same modifications in different neural circuitry, different modifications in the same circuitry, or different modifications in different circuitry. We typically cannot distinguish among these three types of differences at the behavioral level. Therefore we use the phrase *different neural processes* to refer to all three possibilities.

Learning on the Trained Condition One indication that training on different tasks affects different neural processes is that the pattern of learning on the trained condition differs across tasks. The specific assumption here is that the pattern of learning during the training itself differs across different trained conditions only if different neural processes are affected by the training. Such differences are evident in several aspects of learning on the trained condition, including the time frame of learning, the shape of the learning curve, the pattern of performance within each training session, and the amount of daily training required to yield improvement across sessions.

First, the time frame of learning differs markedly across different tasks. In perhaps the clearest demonstration of such differences, we trained five different groups of listeners each on a different basic auditory discrimination task, using similar training and testing regimens in all five cases. In addition, in three of the five cases, we used the same standard stimulus, against which all discrimination comparisons were made. Despite these similarities, the time frame of learning differed across the tasks (figure 24.1). For the discrimination of the sound intensity of brief tone pips (Wright & Fitzgerald, 2005) and of interaural time differences (a sound localization cue) in longer pure tones (Wright & Fitzgerald, 2001), learning appeared to be complete following an approximately two-hour pretraining test, because listeners who received an additional 6–10 hours of training showed no more improvement than controls who participated only in pretraining and posttraining tests. In contrast, for the discrimination of the frequency of (Wright & Sabin, 2007), and temporal interval between (Wright, Buonomano, Mahncke, & Merzenich, 1997), brief tone pips and of interaural level differences (a sound-localization cue) in longer pure tones (Wright & Fitzgerald, 2001), the listeners who received the multiple-hour training improved more than controls did. In general, learning on auditory tasks appears to be complete after times ranging from less than approximately 20 minutes (e.g., interaural-time-difference discrimination) (Ortiz & Wright, 2009) to more than 20 hours (e.g., learning of tone sequences) (Leek & Watson, 1984, 1988) (for a review, see Watson, 1980). Interestingly, while there is some indication that the amount of training that is needed to reach asymptotic performance on auditory tasks increases with increases in task

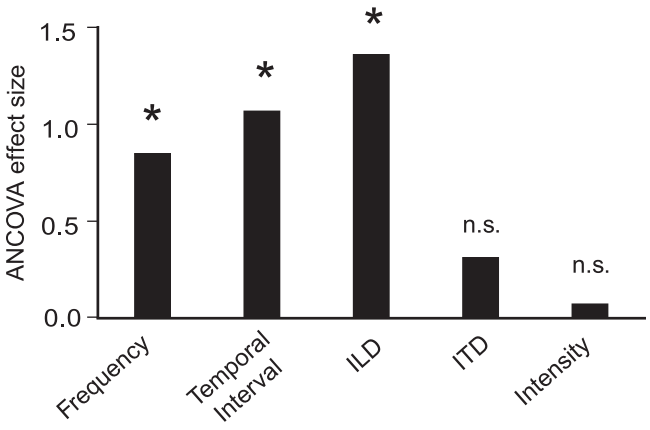


Figure 24.1 Different effects of multiple-hour training on five basic auditory tasks. Each bar shows the effect size of an analysis of covariance conducted on the posttraining thresholds for a given task, with the pretraining thresholds as the covariate. Asterisks indicate that the listeners who practiced 6–10 hours on that task (n = 6–20) had significantly lower posttraining thresholds (better performance) than did controls (n = 6–16) who participated only in pretraining and posttraining tests but received no intervening training; n.s. indicates that the posttraining thresholds did not differ significantly between the groups. Results are shown for frequency, temporal-interval, interaural-level-difference (ILD), interaural-time-difference (ITD), and intensity discrimination.

and stimulus complexity (Watson, 1980), the current cases illustrate that it can differ markedly even among tasks with similar levels of complexity. Rapid improvements on perceptual tasks have been attributed to cognitive processes such as familiarization with the testing procedure and recognition of the sound attribute that is key to performing the task, while more gradual improvements have been attributed to changes in the perceptual systems themselves (e.g., Recanzone, Schreiner, & Merzenich, 1993; Robinson & Summerfield, 1996; Wright & Fitzgerald, 2001). However, this interpretation should be treated with caution because of evidence that perceptual changes can occur quite early in training (e.g., Fiorentini & Berardi, 1980; Ortiz & Wright, 2009; Rubin, Nakayama, & Shapley, 1997).

Second, even when the time frame of learning is similar, in some cases, the shape of the learning curve differs across tasks. For example, of the three tasks described above for which learning continued over multiple daily sessions, the learning curves had different shapes (figure 24.2*A*). For temporal-interval and interaural-level-difference discrimination, the learning rate was most rapid over the first three days of training and slower thereafter. In contrast, for frequency discrimination, the learning rate was approximately constant across the 10 training days. Further, even on the same task, the learning curve sometimes differs in shape for different trained stimuli. For instance, for interaural-level-difference discrimination, learning was initially rapid and then slowed for a pure-tone stimulus but occurred at an

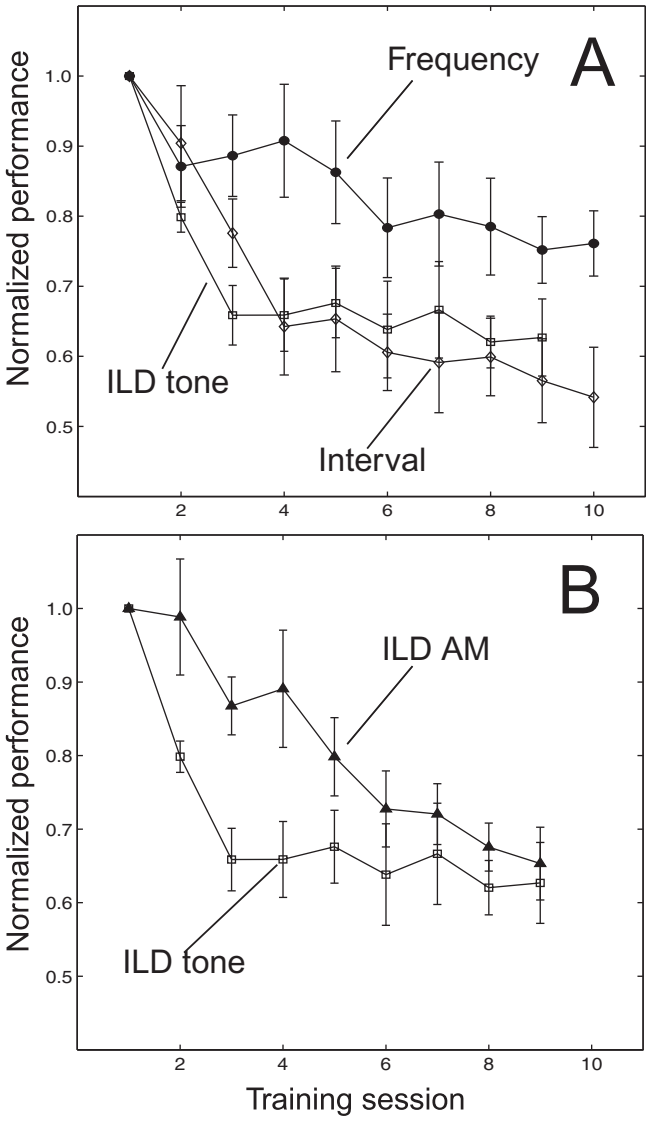


FIGURE 24.2 Different learning-curve shapes for different trained tasks and for different trained stimuli on the same task. Mean thresholds for each training session, normalized as a percentage of the first-session performance, for frequency (*A*, solid circles) and temporal-interval (*A*, open diamonds) discrimination with the same pair of brief tone pips and for interaural-level-difference (ILD) discrimination with a longer pure tone (*A*, *B*, open squares) and an amplitude-modulated (AM) tone (*B*, solid triangles). Each symbol represents the average of 6–10 trained listeners.

approximately constant rate for a sinusoidally amplitude modulated stimulus (figure 24.2*B*). Similarly, for the discrimination of the fundamental frequency of harmonic complexes, the slope of the learning curve was significantly steeper when the individual harmonics could be resolved by the peripheral auditory system than when they could not (Grimault, Micheyl, Carlyon, & Collet, 2002).

Third, in addition to the differences in across-session improvement, the pattern of performance within each training session also differs across tasks. We recently reported that listeners who improved on either temporal-interval or frequency discrimination across multiple daily training sessions showed no systematic improvement in performance within each session (Wright & Sabin, 2007). That is, the improvement appeared to occur between, rather than within, sessions. However, we have seen different within-session patterns for other tasks. For example, although listeners improved across sessions on the detection of a brief tone that was presented immediately before a masking noise (backward masking), their performance tended to worsen from the beginning to the end of each training session (unpublished data). That the within-session performance patterns differ across tasks, despite across-session improvements in all cases, is consistent with the idea that there are at least two distinct stages of perceptual learning. One, acquisition, is the period during which the task is actually practiced. The other, consolidation, is the period during which performance stabilizes or further improves without additional practice, presumably through the transfer of what has been learned during acquisition from short- to long-term memory. The different within-session patterns suggest that the mechanisms that are involved in the acquisition stage are task dependent but nevertheless all are able to lead to consolidation.

Fourth, a related observation is that the amount of training in each daily session that is required to yield improvement across sessions differs for different tasks and possibly even for different stimuli (Wright & Sabin, 2007). We trained four groups of listeners, two on frequency discrimination (figure 24.3*A*) and two on temporal-interval discrimination (figure 24.3*C*), each for six days, with the same standard stimulus. For each task, one group was trained for 360 trials per day, and the other group was trained for 900 trials per day. The listeners who were trained on temporal-interval discrimination showed similar improvements regardless of the amount of daily training (figure 24.3*D*). In contrast, for frequency discrimination, only the listeners who were trained for 900 trials per day improved (figure 24.3*B*). Thus at least for the particular standard stimulus that we used, learning on frequency discrimination required more trials of training per day than did learning on temporal-interval discrimination. Note, however, that performance on frequency discrimination with a different standard stimulus improved over eight 350-trial sessions (Roth, Amir, Alaluf, Buchsenspanner, & Kishon-Rabin, 2003), suggesting that the amount of daily training required for learning may be stimulus dependent as well as task dependent (Wright & Sabin, 2007). In the context of the two learning stages, these results suggest that different amounts of training during acquisition are required to trigger consolidation for different tasks and even for different stimuli.

ACROSS-TASK GENERALIZATION Another line of evidence that auditory training affects different neural processes for

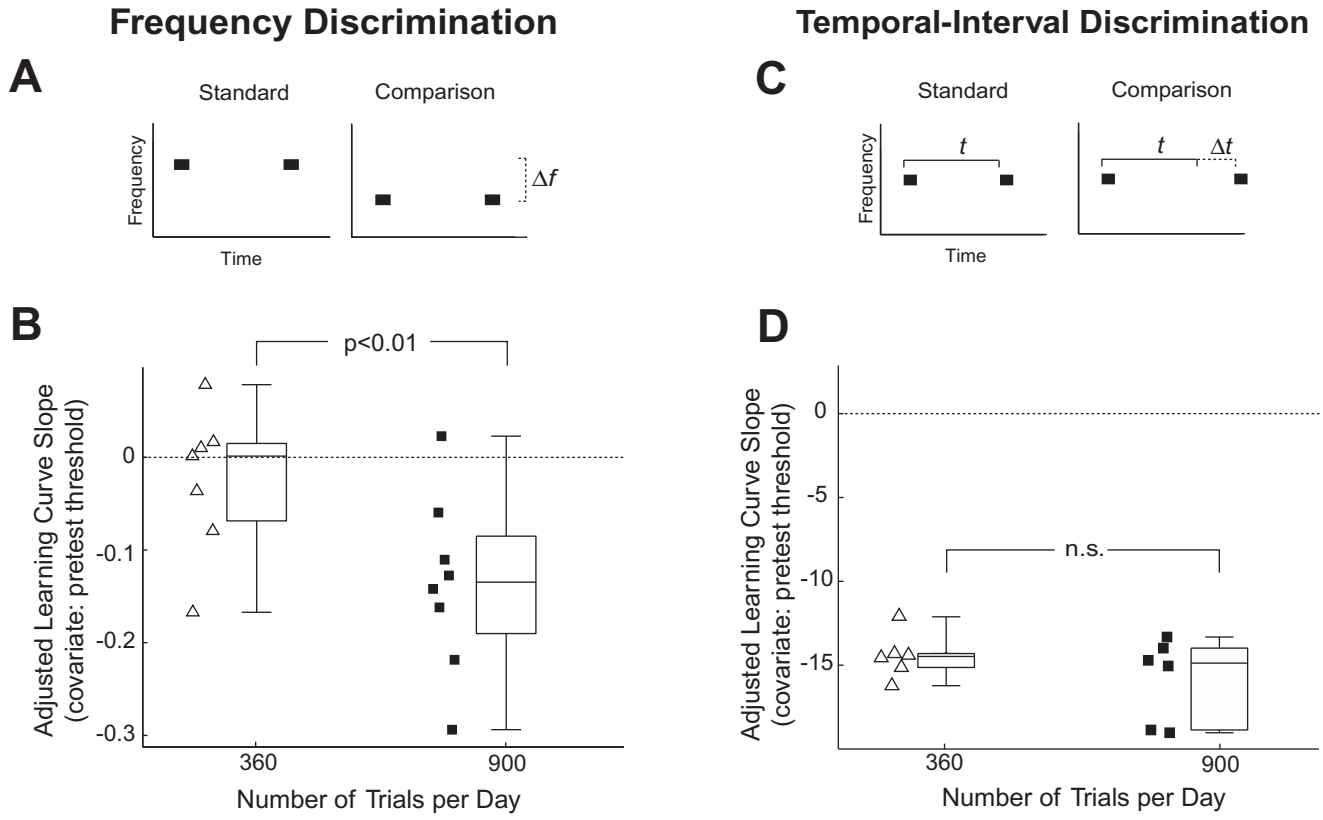


FIGURE 24.3 Different amounts of daily training required for learning across multiple sessions on frequency (*A*, *B*) and temporal-interval (*C*, *D*) discrimination. (*A*) Schematic diagram of the frequency discrimination task. Listeners discriminated between standard (left) and comparison (right) stimuli that differed from each other only in frequency. (*B*) Rate of across-session improvement on frequency discrimination indicated by the slopes of regression lines fitted, for each listener (symbols), to the daily thresholds versus the log of the training session number. Individual differences were taken into account by adjusting the slopes based on pretraining thresholds (ANCOVA). The box plots indicate the median and quartile values. The slopes did not differ significantly from zero for listeners who practiced 360 trials per day for six days (open triangles; $n = 7$), indicating no improvement across training sessions. In contrast, the slopes of listeners who practiced 900 trials per day (solid squares; $n = 8$) differed significantly from zero and were negative ($p < 0.01$), indicating across-session improvement. (*C*, *D*) Same as *A* and *B* but for the temporal-interval discrimination task. For this task, the slopes were significantly different from zero and were negative, regardless of whether the listeners practiced 360 (open triangles; $n = 6$, $p < 0.001$) or 900 (solid squares; $n = 6$, $p < 0.0001$) trials per day for six days, indicating improvement across training sessions in both cases. (Figure adapted from Wright & Sabin, 2007.)

different trained tasks is that training on one task rarely leads to performance improvements on other tasks (Wright & Zhang, 2009). The specific assumption here is that learning generalizes from a trained to an untrained task if and only if the practice on the trained task modifies neural processes that also govern performance on the untrained one. Therefore a failure to generalize across tasks suggests that different neural processes are modified through training on those tasks.

There are a number of examples of a lack of across-task generalization on basic auditory skills. Following multiple-session training, learning did not generalize in either direction between frequency and amplitude-modulation rate discrimination (figure 24.4) (Fitzgerald & Wright, 2005; Grimault, Micheyl, Carlyon, Bacon, & Collet, 2003), asynchrony detection and order discrimination at sound onset (Mossbridge, Fitzgerald, O'Connor, & Wright, 2006), or frequency and temporal-interval discrimination (unpublished data). It also did not generalize, in the one direction that was tested, from interaural-level-difference to interaural-time-difference discrimination (Wright & Fitzgerald, 2001), from amplitude-modulation rate to rippled-noise (figure 24.4) (Fitzgerald & Wright, 2005) or temporal-interval (van Wassenhove & Nagarajan, 2007) discrimination, or from amplitude-modulation rate discrimination to amplitude-modulation detection (figure 24.4) (Fitzgerald & Wright, 2005). The lack of across-task generalization has also been observed following a single session of training. For example, a brief period of training on sound-intensity or visual-contrast discrimination did not lead to better performance on frequency discrimination, though the same period of training on frequency discrimination did (Hawkey, Amitay,

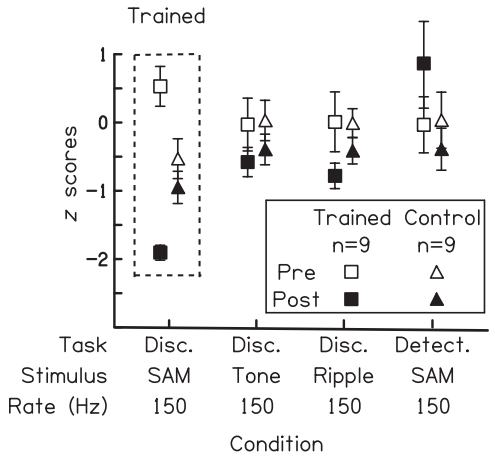


FIGURE 24.4 Lack of across-task generalization following learning on amplitude-modulation rate discrimination. The mean threshold values on a set of conditions tested before and after nine 720-trial daily training sessions on a single amplitude modulation rate discrimination condition. Trained listeners ($n = 9$; squares) improved significantly more than controls did (triangles) between the pretraining ($n = 9$; open symbols) and posttraining (solid symbols) tests on the trained condition (left column) but not on pure-tone frequency discrimination, rippled-noise discrimination, or amplitude modulation detection. The parameters for each condition are marked on the abscissas. The box indicates significantly more improvement in the trained listeners than in controls ($p < 0.05$). (Figure adapted from Fitzgerald & Wright, 2005.)

& Moore, 2004). Further, though listeners who received a single session of practice on temporal-interval or interaural-level-difference discrimination showed better performance on interaural-time-difference discrimination on the following day than naïve listeners did, this improvement was significantly smaller than that shown by listeners who were trained on interaural-time-difference discrimination itself (Ortiz & Wright, 2009). The need for practice on the target task to obtain learning on that task has been taken as evidence that the particular neural processes that are to be modified by training are selected via top-down influences, for example, through attention to the particular sound attribute that is key to performing the target task (Hochstein & Ahissar, 2002).

It should be noted that learning occasionally does generalize across tasks, but so far, the across-task generalization has been observed only in one direction. To date, the clearest instance of such unidirectional generalization is that learning generalized from asynchrony detection to order discrimination at sound offset but not vice versa (Mossbridge, Scissors, & Wright, 2008) (see the section "Relative-Timing Tasks" below). Similarly, a comparison between two other investigations suggests that learning on pure-tone frequency discrimination generalized to fundamental-frequency discrimination (Grimault et al., 2002) but not the reverse (Demany & Semal, 2002). In these cases, training on the different tasks might have engaged different but related neural processes.

ACROSS-STIMULUS GENERALIZATION A third line of evidence that practice on different auditory tasks influences different neural processes is that the pattern of generalization to the trained task with untrained stimuli differs across tasks. Similar to the assumption about across-task generalization, the assumption here is that learning on a trained task generalizes from a trained stimulus to an untrained one if and only if training modifies neural processes that govern performance with both stimuli. Therefore different patterns of across-stimulus generalization for different trained tasks indicate that the training affected neural processes with different tuning characteristics for each task.

Differences in the across-stimulus generalization patterns are apparent in two forms of comparison across tasks. First, the generalization results for a given sound attribute differ across tasks. Across-stimulus generalization to three different stimulus attributes has been examined on multiple tasks. These attributes are frequency, timing, and modality. For each of these attributes, learning generalized to stimuli with untrained values of that attribute for some tasks but not for others. Generalization across frequencies has been reported for pure-tone frequency (Amitay, Hawkey, & Moore, 2005; Delhommeau, Micheyl, & Jouvent, 2005; Demany & Semal, 2002; Irvine, Martin, Klimkeit, & Smith, 2000) and temporal-interval (Karmarkar & Buonomano, 2003; Wright et al., 1997) discrimination. However, learning was specific to the trained frequency for interaural-level-difference discrimination with pure tones (Wright & Fitzgerald, 2001) and tone detection in quiet (Zwislocki, Maire, Feldman, & Rubin, 1958) and to the trained tone pair for asynchrony detection and order discrimination (Mossbridge et al., 2006; Mossbridge, Scissors, & Wright, 2008). In terms of timing, learning generalized at least partially to stimuli with untrained durations (Delhommeau et al., 2005) or temporal intervals (Wright & Fitzgerald, 2005) for frequency discrimination but was specific to the trained interval for temporal-interval discrimination (Karmarkar & Buonomano, 2003; Wright et al., 1997) and for the detection of short tones in same-duration gated noise (Tucker, Williams, & Jeffress, 1968). Finally, for modality, learning on temporal-interval discrimination generalized from the auditory system to motor performance (Meegan, Aslin, & Jacobs, 2000) and from the somatosensory to the auditory system (Nagarajan, Blake, Wright, Byl, & Merzenich, 1998), but there is some indication that learning on asynchrony detection does not generalize across modalities (Virsu, Oksanen-Hennah, Vedenpaa, Jaatinen, & Lahti-Nuuttila, 2008). Second, the across-stimulus generalization pattern differs across tasks when the comparison is made relative to the respective trained sound attribute of those tasks. Learning generalized across different values of the *trained* attribute for some tasks but not others. For example, learning generalized to untrained frequencies for frequency discrimination (Amitay

et al., 2005; Delhommeau et al., 2005; Demany & Semal, 2002; Irvine et al., 2000) and to an untrained standard interaural-level-difference value for interaural-level-difference discrimination (Wright & Fitzgerald, 2001) but not to untrained temporal intervals for temporal-interval discrimination (Karmarkar & Buonomano, 2003; Wright et al., 1997). The differences in across-stimulus generalization pattern for different trained tasks have been used to make inferences about the tuning characteristics of the neural circuitry that was modified by training (see below).

Characteristics of auditory processing revealed by perceptual-learning patterns

Here, we provide a systematic examination of the effect of perceptual training on each of two selected aspects of auditory perception, and the characteristics of the neural processing that can be inferred from those results. These inferences are made based on the assumptions described in the preceding sections. Note that these inferences apply only to the subset of neural processes that was affected by the training, though many may be potentially engaged while performing the trained condition.

Temporal Tasks A critical role of the auditory system is to encode the duration of and the temporal relationship between events. Here, we present two sets of experiments in which we used perceptual learning patterns to gain insights into the neural processes underlying these auditory-timing abilities.

Temporal-interval discrimination In one set of experiments, we examined learning on a temporal-interval discrimination task. In the first experiment (Wright et al., 1997), we trained listeners to discriminate deviations from a standard temporal interval of 100 ms, marked by two brief 1-kHz tone pips, during multiple daily sessions. Nearly all of these listeners showed significant learning during the training phase. This learning generalized to an untrained condition that differed from the trained one only in the frequency of the stimulus (4 kHz versus 1 kHz) but did not generalize to untrained conditions that differed from the trained one only in the duration of the temporal interval (50, 200, or 500 ms versus 100 ms) (figure 24.5). After the original experiment, we administered to a control group the pretraining and posttraining tests but not the multiple-day training. The controls did not show any improvement, and a between-group comparison of the trained listeners and controls yielded the same conclusions as were reached for the trained listeners alone (Wright & Fitzgerald, 2005). Thus the learning on auditory temporal-interval discrimination was specific to the trained interval but generalized across stimulus frequency. Karmarkar and Buonomano (2003) replicated this generalization pattern and also reported generalization to an untrained marker type (a continuous tone versus tone pips) at the trained interval.

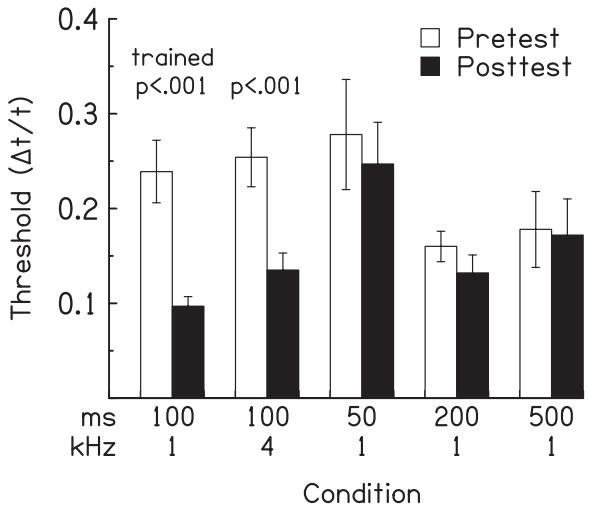


Figure 24.5 Learning and generalization on auditory temporal-interval discrimination. The mean threshold values on five temporal-interval discrimination conditions tested before (open bars) and after (solid bars) ten 900-trial daily training sessions on a single temporal-interval discrimination condition. The conditions are marked on the abscissa by the temporal interval (in milliseconds) and the tone frequency (in kilohertz) of the standard stimulus. Thresholds differed significantly between the pretraining and posttraining tests for the trained condition (100 ms at 1 kHz), indicating that training led to learning. For the remaining conditions, the thresholds differed significantly between the pretraining and posttraining tests only for the untrained frequency (100 ms at 4 kHz) but not for the untrained temporal intervals (50, 200, or 500 ms at 1 kHz). Thus the learning generalized across frequency but was specific to the trained temporal interval. Results are shown only for listeners who improved significantly across the training sessions ($n = 11$ out of 14 tested for three conditions: 100 and 200 ms at 1 kHz and 100 ms at 4 kHz; $n = 5$ out of 6 tested for the remaining two conditions: 50 and 500 ms at 1 kHz). (Figure from Wright, Buonomano, Mahncke, & Merzenich, 1997.)

We subsequently trained another group of observers on temporal-interval discrimination in the somatosensory system (Nagarajan et al., 1998). Parallel to the pattern observed in the auditory system, trained observers improved their performance on the trained condition, and this learning generalized to untrained positions on the trained hand and to the untrained hand but was largely specific to the trained temporal interval. Most interestingly, training in the somatosensory system generalized to the auditory system but only for the trained interval. Interval-specific generalization of interval-discrimination learning across systems has also been observed from auditory to motor performance (a motor tapping task) (Meegan et al., 2000).

Taken together, these results indicate that the neural modifications induced by the training paradigms that were used in these experiments influence auditory and somato-

sensory as well as motor performance on temporal-interval discrimination with a variety of marker conditions but only for a specific interval. These characteristics reveal a neural process that encodes temporal intervals in interval-specific channels and that is possibly located beyond the primary sensory or motor cortices.

Relative-timing tasks In another set of experiments, we examined learning on two relative-timing tasks: asynchrony detection and order discrimination. In an asynchrony-detection task, the object is to determine whether different components of a complex sound start or end at the same time (synchronously) or at different times (asynchronously). This ability helps listeners to separate multiple auditory events. For order-discrimination, the task is to determine the order in which different components of a complex sound start or end. This ability helps listeners to distinguish, for example, words (e.g., "pat" and "tap") and musical melodies (e.g., ascending versus descending scales) perception.

In the first experiment (Mossbridge et al., 2006), we trained one group of listeners on asynchrony detection and another group on order discrimination using standard stimuli composed of 0.25- and 4-kHz tones, with the crucial information presented at sound onset (figure 24.6*A*–*C*). We used the same multiple-hour training paradigm for both groups. Controls, who did not receive the training, improved between the pretraining and posttraining tests on the trained conditions as well as on a variety of related untrained conditions, indicating that exposure to the pretraining test itself led to learning. However, both trained groups improved more than controls did on their respective trained conditions, demonstrating that the multiple-hour training induced additional learning beyond that resulting from exposure to the pretraining test. This training-induced learning did not generalize to untrained tone pairs (e.g., 0.5 and 1.5, 0.75 and 1.25 versus 0.25 and 4 kHz), to an untrained temporal position (sound offset versus onset), or to the other, untrained, task (asynchrony versus order or vice versa).

We later trained two new groups of listeners on the same tasks at sound offset (Mossbridge et al., 2008), using the same standard stimuli as in the original investigation (figures 24.6*A*, 24.6*D*, and 24.6*E*). Unlike at sound onset, at sound offset the control listeners in large part showed no improvement between the pretraining and posttraining tests. Also different from sound onset, and more important in the present context, the generalization patterns differed between the asynchrony and order tasks. Both groups who were trained at sound offset improved more than controls did on their respective trained conditions, demonstrating training-induced learning. For order discrimination at sound offset (figure 24.6*E*), as for both tasks at sound onset (figures 24.6*B* and 24.6*C*), the training-induced learning did not generalize to untrained tone pairs (0.5 and 1.5, 0.75 and 1.25 versus 0.25 and 4 kHz), to an untrained temporal position (sound onset versus offset), or to the untrained task (asynchrony versus order). However, for asynchrony detection at sound offset (figure 24.6*D*), while the training-induced learning was still specific to the trained tone pair, it did generalize both to the untrained temporal position (sound onset) and to the untrained task (order).

These patterns of learning and generalization provide insights into the neural processing of relative-timing tasks (Mossbridge et al., 2006, 2008). That performance improved with training for both asynchrony detection and order discrimination at both sound onset and offset indicates that the neural processes underlying relative-timing judgments are malleable. More interestingly, the lack of mutual generalization across tasks and across temporal positions suggests that the training on the four conditions affected different neural processes. However, these neural processes appear to be related, as is suggested by the unidirectional generalization of learning from asynchrony detection at sound offset to the remaining three cases. One possible form of this relationship is that performance on each task, at each temporal position, is governed by a specified neural process and that there is a unidirectional dependency among these processes. Another possibility is that while performance on asynchrony detection at sound offset is governed by a general relative-timing mechanism, performance on the other three conditions can be affected by both the global mechanism and the condition-specific mechanisms and that the training on those three conditions modified the specific ones. Regardless of the relationship, the neural processes underlying learning in all four training groups appear to be tuned to specific frequency pairs.

Spatial Tasks Another key role of the auditory system is to encode the spatial location of sound sources. The two primary cues to sound-source position on the horizontal plane are interaural level differences (ILDs) and interaural time differences (ITDs). These cues arise because the sound from a given source can reach the two ears at different times or with different levels, depending on the frequency content and position of the source. We have examined the effect of training on the ability of listeners to detect small variations in ILDs and ITDs in a series of experiments. We use the results to make inferences about the neural processing of these two sound localization cues.

In our first investigation of this issue (Wright & Fitzgerald, 2001), we trained two groups of listeners for multiple sessions: one on ILD discrimination with a high-frequency (4 kHz) tone and the other on ITD discrimination with a low-frequency (0.5 kHz) tone. We used different stimulus frequencies for the two cues because ILDs are known to be most effective at high frequencies and ITDs are known to be most effective at low frequencies. To manipulate the two

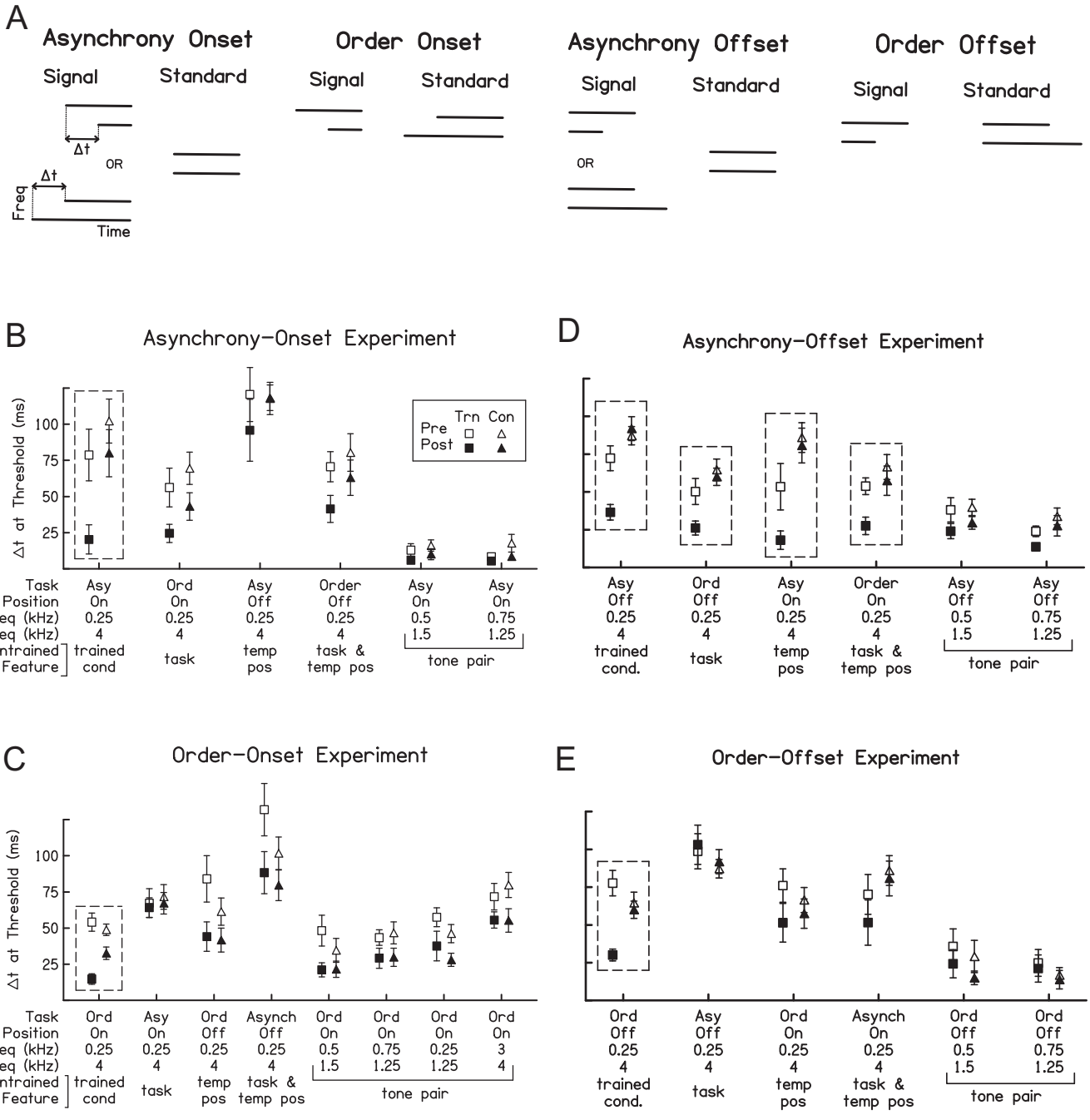


FIGURE 24.6 Learning and generalization on auditory asynchrony detection (*B*, *D*) and temporal-order discrimination (*C*, *E*). (*A*) Schematic diagrams of the signal and standard stimuli used in the four relative-timing conditions. Each stimulus consisted of two tones. The duration of the higher-frequency tone was fixed at 500 ms. The frequencies of the tones depended on the condition parameters. (*B*–*E*) The mean threshold values on a set of relative-timing conditions tested before and after six to eight 720-trial daily training sessions on a single condition: asynchrony-detection (*B*) or temporal-order-discrimination (*C*) at sound onset or asynchrony-detection (*D*) or temporal-order-discrimination (*E*) at sound offset. In all four cases, trained listeners (squares) improved significantly more than controls (triangles) between the pretraining (open symbols) and posttraining (filled symbols) tests on the trained condition (left column). However, the generalization pattern differed across the trained conditions. In three cases (*B*, *C*, *E*), the learning attributable to the multiple-hour training was specific to the trained condition. In the fourth case (*D*), the learning generalized to all conditions tested with the trained frequency pair and therefore spread more broadly than learning for the other trained conditions did. The parameters for each condition are marked on the abscissas (n = 6–18 for each group in each condition). Boxes indicate conditions on which trained listeners learned significantly more than controls did ($p < 0.05$). (Figure adapted from Mossbridge, Fitzgerald, O'Connor, & Wright, 2006; Mossbridge, Scissors, & Wright, 2008.)

cues separately, we presented the sounds over headphones. The listener's task was to discriminate changes in the lateral position of the sound image that were caused by changes in one of the two cues; therefore the task is referred to as ILD or ITD discrimination depending on the cue manipulated. The multiple-hour training had markedly different effects on ILD and ITD discrimination (figure 24.7) (Wright & Fitzgerald, 2001). Control listeners, who participated only in approximately 2-hour pretraining and posttraining tests, improved on both cues, indicating that exposure to the pretraining test itself contributed to improvements on these tasks. However, while the listeners who received multiple-hour training on ILD discrimination improved more than the controls did, the listeners who were trained on ITD discrimination did not. Thus under this training regimen, performance on ILD but not ITD discrimination continued to improve during the multiple hours of training following the pretest. In addition, the training-induced learning on ILD discrimination did not affect ITD discrimination. The different time frames of learning on ILD and ITD discrimination and the lack of influence of ILD training on ITD performance suggest differential neural processing of high-frequency ILDs and low-frequency ITDs. Further, the training-induced learning on ILD discrimination generalized to ILD discrimination with an untrained standard location (off-center versus midline) but not to untrained sound frequencies (6 and 0.5 kHz versus 4 kHz). This pattern of across-stimulus generalization suggests that the multihour training on ILD discrimination modified neural circuitry that encodes sounds of a specific frequency but from a variety of locations.

In a follow-up experiment, we tested whether the observed difference between the learning patterns on ILD and ITD discrimination resulted from the difference between the two cues (ILD versus ITD) or from the difference in stimulus frequency (4 kHz versus 0.5 kHz) (Zhang & Wright, 2007). To do so, we trained another group of listeners on ITD discrimination at the same high frequency as the one we had previously used for ILD-discrimination training (4 kHz). However, because at high sound frequencies (>1.5 kHz) humans are not sensitive to ITDs in pure tones, we amplitude-modulated the 4-kHz tone with a sinusoid of 0.3 kHz (a SAM tone), a stimulus that had been reported to yield ITD sensitivity (Henning & Ashton, 1981). In this investigation, the controls improved between the pretraining and posttraining tests, and the trained listeners did not improve more than the controls did. This learning pattern is similar to what we had observed for ITD discrimination at 0.5 kHz but different from that for ILD discrimination at 4 kHz, thereby ruling out the contribution of stimulus frequency to the observed difference in learning of ILD and ITD discrimination.

Most recently, we tested the possibility that the difference in the training results for ILD and ITD discrimination

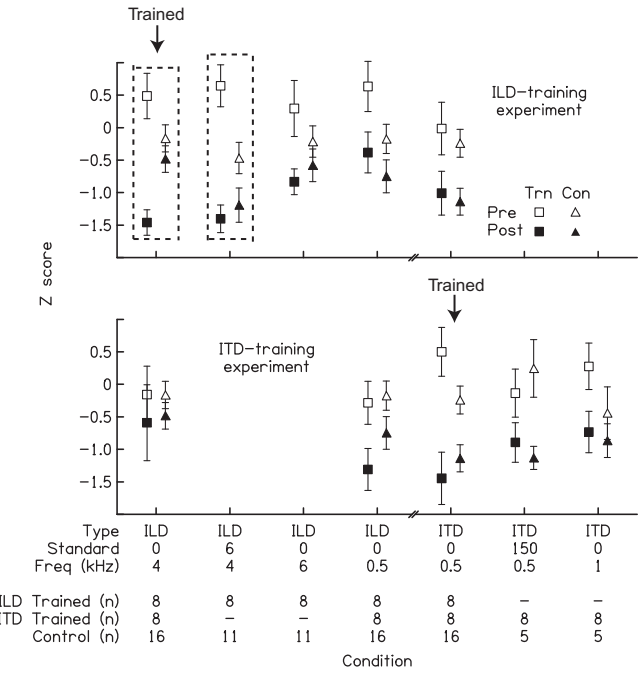


Figure 24.7 Learning and generalization on auditory interaural-level-difference (ILD) or interaural-time-difference (ITD) discrimination with pure tones. The mean threshold values, expressed as *z*-scores, on a set of conditions tested before and after nine 720-trial daily sessions on a single condition, either interaural-level-difference (top) or interaural-time-difference (bottom) discrimination. The parameters and number of listeners for each condition are marked on the abscissa. Trained listeners (squares) who practiced ILD discrimination improved significantly more than controls (triangles) between the pretraining (open symbols) and posttraining (solid symbols) tests on the trained condition (top, left column). This learning generalized to an untrained condition with a different standard ILD value (6 dB) but not to any other untrained condition. In contrast, for the ITD-training experiment, both the trained listeners and controls improved, and there was no between-group difference in any condition tested, either trained (bottom, fifth column) or untrained. Boxes indicate significant differences in improvement between trained listeners and controls ($p < 0.05$). (Figure adapted from Wright & Fitzgerald, 2001.)

at 4 kHz was due to the different stimulus types (pure versus SAM tones) rather than the different cues (ILD versus ITD) (Zhang & Wright, in review). In this experiment, we trained a new group of listeners on ILD discrimination with the same SAM tone that we had used to train high-frequency ITD discrimination. Once again, controls improved between the pretraining and posttraining tests. However, the majority of the trained listeners improved more than controls. Thus multihour training induced additional learning in ILD, but not ITD, discrimination, even for the same stimulus. These results suggest differential plasticity in the processing of the two cues regardless of stimulus frequency or type.

Notably, the detailed pattern of the training results for ILD discrimination also differed between the pure tone and

the SAM tone (Zhang & Wright, in review). There were three such differences. First, the training-induced learning on ILD discrimination with the SAM tone generalized to untrained SAM tones with different carrier frequencies and modulation rates but not to pure tones, even when those tones had the same frequency as the trained carrier or modulation rate. Thus within the trained stimulus type, learning for the SAM tone generalized across frequency, while that for the pure tone did not. Second, the amount of learning could be predicted on the basis of the starting thresholds for ILD learning with the pure tone but not with the SAM tone. Third, the learning curve was more linear for the SAM tone than for the pure tone. These differences suggest that training affects differentially the processing of ILDs in amplitude-modulated stimuli and pure tones, even at the same frequency.

Finally, we investigated the extent to which the rapid improvement that we observed on ITD discrimination results from learning of the trained stimulus, the lateralization task, or other factors that are collectively classified as the procedure (Ortiz & Wright, 2009). Toward this end, we trained three groups of listeners for a single session, each on a different condition, and tested all of them the next day on a target ITD-discrimination condition. The three trained conditions shared different elements with the target ITD condition, forming a hierarchy of similarity. One group of listeners was trained on a temporal-interval discrimination condition that shared with the target condition only the general, procedural aspects. These listeners had lower thresholds on the target ITD condition than naïve listeners did, suggesting procedure learning. Another group of listeners was trained on an ILD-discrimination condition that shared with the target condition both the procedure and the lateralization task but not the stimulus. The ITD-discrimination thresholds of the ILD-trained listeners were similar to those of the interval-trained listeners, implying that there was little additional improvement that was attributable to task learning. The third group was trained on the target ITD condition itself. These listeners had lower ITD thresholds than the ILD-trained listeners did, suggesting that the additional improvement resulted from stimulus learning. Thus rapid improvements on ITD discrimination appear to result primarily from learning of the procedure and the stimulus, implying that a single session of training can affect at least two types of neural processes.

Neural underpinnings of perceptual learning

Up to this point, we have documented the large variation in learning patterns across auditory tasks, argued that this variation suggests that different neural processes are involved in learning on these tasks, and illustrated how these learning patterns can be used to make inferences about the affected processes. Here we speculate about the actual neural underpinnings of auditory learning on the basis of behavioral and physiological data from the auditory as well as other sensory systems.

We propose that in most cases, for auditory learning to occur on a given condition, a neural process that limits the performance on that condition has to be selected and placed in a modification-prone state (sensitized). Sufficient stimulation of the sensitized process results in modifications that lead to behavioral improvement. We further suggest that the selection and sensitization of the targeted process occurs through top-down influences such as attention or reward. These influences are typically and optimally provided by performance of the target condition rather than simply through the bottom-up stimulation received from stimulus exposures. A role for top-down influences in perceptual learning has been proposed previously for visual learning (Ahissar & Hochstein, 2004; Gilbert & Sigman, 2007; Seitz & Watanabe, 2005). The primary behavioral evidence for this involvement, both here and in other sensory systems, comes from the observations that learning on one task rarely generalizes to other tasks, even when the same stimuli are employed, and from the different learning and generalization patterns for different tasks performed with the same stimuli. The idea that sufficient stimulation of the sensitized process is required to achieve learning comes in part from the demonstration that improvement across days on an auditory task requires a sufficient amount of training per day (Wright & Sabin, 2007). It also echoes a recent proposal, arising from a literature review, that a "learning threshold" must be surpassed, through any of a variety of means, for improvement to occur (Seitz & Dinse, 2007). Note that this proposed requirement for learning provides one means for preserving the necessary balance between stability and plasticity in the nervous system.

We also suggest that the processes that are selected and sensitized during auditory training differ across tasks and can shift over the course of training. These ideas are supported by evidence from neurophysiology and imaging that the neural changes that accompany perceptual learning occur at multiple stages of the nervous system, including primary sensory cortices (Clapp, Kirk, Hamm, Shepherd, & Teyler, 2005; Furmanski, Schluppeck, & Engel, 2004; Li, Piech, & Gilbert, 2008; Pourtois, Rauss, Vuilleumier, & Schwartz, 2008) as well as associative (Law & Gold, 2008) and frontal (Krigolson, Pierce, Holroyd, & Tanaka, 2008) cortices, particularly those involved in attention (Mukai et al., 2007). There are also reports of global reorganization spanning multiple stages of processing (Schiltz, Bodart, Michel, & Crommelinck, 2001; Sigman et al., 2005; Vaina, Belliveau, des Roziers, & Zeffiro, 1998; van Wassenhove & Nagarajan, 2007). These ideas receive further support from evidence that different sites are affected at different time

points in training (Atienza, Cantero, & Dominguez-Marin, 2002; Gottselig, Brandeis, Hofer-Tinguely, Borbely, & Achermann, 2004; Karni et al., 1998; Petersen, van Mier, Fiez, & Raichle, 1998) and that changes in primary cortex that occur during learning can reverse after learning is complete, leaving the cortex in its original state (Yotsumoto, Watanabe, & Sasaki, 2008). Behavioral data have also been used to argue that a variety of different sites are affected by perceptual training. These arguments are based on the practice of matching generalization patterns to the tuning functions at different processing levels. For example, in the visual system, the specificity of learning to basic stimulus attributes has been taken as evidence for modifications in early stages of the visual system, because the neurons at those stages are tuned to those attributes (Ahissar & Hochstein, 2004; Fahle, 2004; Karni & Sagi, 1991; Poggio, Fahle, & Edelman, 1992). Similarly, broader generalization has been attributed to modifications in later visual-processing stages in which neural tuning is also less selective (Ahissar & Hochstein, 2004). However, it has been noted that both generalization patterns also could result from modifications in a common central site that interprets the sensory information obtained from earlier processing stages, with the degree of specificity reflecting the reweighting of sensory information based on different task demands (Mollon & Danilova, 1996). As to which site will be modified, one proposal is that training affects the most central level of representation that provides an adequate signal-to-noise ratio for task performance (Ahissar & Hochstein, 2004). Thus the site of modification is determined by task difficulty, with difficult tasks affecting more peripheral sites. This idea is based primarily on the observation that learning on an easy visual task generalized more broadly than that on a more difficult task (Ahissar & Hochstein, 2004).

Finally, we think that auditory learning could be mediated by different types of neural changes for different tasks. In support of this view, perceptual learning has been associated with a variety of physiological changes, including possibly interrelated changes such as the remapping of stimulus representations (Feldman & Brecht, 2005; Recanzone, Merzenich, Jenkins, Grajski, & Dinse, 1992), increased (Furmanski et al., 2004; van Wassenhove & Nagarajan, 2007) or decreased (Sigman et al., 2005; Vaina et al., 1998) neural activity, inhibition of irrelevant channels (Casco, Campana, Grieco, & Fuggetta, 2004), sharper (Schoups, Vogels, Qian, & Orban, 2001) or broader (Crist, Li, & Gilbert, 2001) neural tuning, improved neural response reliability (Yao, Shi, Han, Gao, & Dan, 2007), and alteration in synaptic (Yao & Dan, 2005) and cellular (Barkai, 2005) properties of neurons (for a review, see Buonomano & Merzenich, 1998). These changes differ in both their time course and their range of influence. Additional evidence for the involvement of different mechanisms comes from mathematical models, which, when fitted to behavioral data, suggest that in some cases, perceptual learning results from an enhancement of the signal representation, while in others, it results from a reduction in noise (Dosher & Lu, 1998; Gold, Bennett, & Sekuler, 1999; Lu, Chu, Dosher, & Lee, 2005).

Conclusion

Training induces improvements in a variety of basic auditory skills in humans but does so with different dynamics and generalization patterns. These differences support the idea that perceptual training on different skills affects different neural processes. Detailed examinations of the patterns of learning and generalization provide information about the characteristics of the neural processes that are affected by the training. Thus perceptual training can be used as a noninvasive tool to probe into the neural substrates of sound perception. A greater understanding of the rules and mechanisms of auditory learning can also guide the development of training regimens to improve auditory skills.

REFERENCES

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.*, *8*(10), 457–464.

Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Percept. Psychophys.*, *67*(4), 691–698.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learn. Mem.*, *9*(3), 138–150.

Barkai, E. (2005). Dynamics of learning-induced cellular modifications in the cortex. *Biol. Cybern.*, *92*(6), 360–366.

Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: From synapses to maps. *Annu. Rev. Neurosci.*, *21*, 149–186.

Casco, C., Campana, G., Grieco, A., & Fuggetta, G. (2004). Perceptual learning modulates electrophysiological and psychophysical response to visual texture segmentation in humans. *Neurosci. Lett.*, *371*(1), 18–23.

Clapp, W. C., Kirk, I. J., Hamm, J. P., Shepherd, D., & Teyler, T. J. (2005). Induction of LTP in the human auditory cortex by sensory stimulation. *Eur. J. Neurosci.*, *22*(5), 1135–1140.

Crist, R. E., Li, W., & Gilbert, C. D. (2001). Learning to see: Experience and attention in primary visual cortex. *Nat. Neurosci.*, *4*(5), 519–525.

Delhommeau, K., Micheyl, C., & Jouvent, R. (2005). Generalization of frequency discrimination learning across frequencies and ears: Implications for underlying neural mechanisms in humans. *J. Assoc. Res. Otolaryngol.*, *6*(2), 171–179.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *J. Acoust. Soc. Am.*, *111*(3), 1377–1388.

Dosher, B. A., & Lu, Z. L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proc. Natl. Acad. Sci. USA*, *95*(23), 13988–13993.

Fahle, M. (2004). Perceptual learning: A case for early selection. *J. Vis.*, *4*(10), 879–890.

Feldman, D. E., & Brecht, M. (2005). Map plasticity in somatosensory cortex. *Science*, *310*(5749), 810–815.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*(5777), 43–44.
Fitzgerald, M. B., & Wright, B. A. (2005). A perceptual learning investigation of the pitch elicited by amplitude-modulated noise. *J. Acoust. Soc. Am.*, *118*(6), 3794–3803.
Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Curr. Biol.*, *14*(7), 573–578.
Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*(5), 677–696.
Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, *402*(6758), 176–178.
Gottselig, J. M., Brandeis, D., Hofer-Tinguely, G., Borbely, A. A., & Achermann, P. (2004). Human central auditory plasticity associated with tone sequence learning. *Learn. Mem.*, *11*(2), 162–171.
Grimault, N., Micheyl, C., Carlyon, R. P., Bacon, S. P., & Collet, L. (2003). Learning in discrimination of frequency or modulation rate: Generalization to fundamental frequency discrimination. *Hear. Res.*, *184*(1–2), 41–50.
Grimault, N., Micheyl, C., Carlyon, R. P., & Collet, L. (2002). Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. *Percept. Psychophys.*, *64*(2), 189–197.
Hawkey, D. J., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nat. Neurosci.*, *7*(10), 1055–1056.
Henning, G. B., & Ashton, J. (1981). The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay. *Hear. Res.*, *4*(2), 185–194.
Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.
Irvine, D. R., Martin, R. L., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *J. Acoust. Soc. Am.*, *108*(6), 2964–2968.
Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learn. Mem.*, *10*(2), 141–147.
Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., et al. (1998). The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proc. Natl. Acad. Sci. USA*, *95*(3), 861–868.
Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proc. Natl. Acad. Sci. USA*, *88*(11), 4966–4970.
Krigolson, O. E., Pierce, L. J., Holroyd, C. B., & Tanaka, J. W. (2008). Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise. *J. Cogn. Neurosci.* [Epub ahead of print. doi: 10.1162/jocn.2009.21128.]
Law, C. T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.*, *11*(4), 505–513.
Leek, M. R., & Watson, C. S. (1984). Learning to detect auditory pattern components. *J. Acoust. Soc. Am.*, *76*(4), 1037–1044.
Leek, M. R., & Watson, C. S. (1988). Auditory perceptual learning of tonal patterns. *Percept. Psychophys.*, *43*(4), 389–394.
Li, W., Piech, V., & Gilbert, C. D. (2008). Learning to link visual contours. *Neuron*, *57*(3), 442–451.
Lu, Z. L., Chu, W., Dosher, B. A., & Lee, S. (2005). Independent perceptual learning in monocular and binocular motion systems. *Proc. Natl. Acad. Sci. USA*, *102*(15), 5624–5629.
Meegan, D. V., Aslin, R. N., & Jacobs, R. A. (2000). Motor timing learned without motor training. *Nat. Neurosci.*, *3*(9), 860–862.
Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vis.*, *10*(1), 51–58.
Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., & Wright, B. A. (2006). Perceptual-learning evidence for separate processing of asynchrony and order tasks. *J. Neurosci.*, *26*(49), 12708–12716.
Mossbridge, J. A., Scissors, B. N., & Wright, B. A. (2008). Learning and generalization on asynchrony and order tasks at sound offset: Implications for underlying neural circuitry. *Learn. Mem.*, *15*(1), 13–20.
Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., & Ungerleider, L. G. (2007). Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *J. Neurosci.*, *27*(42), 11401–11411.
Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., & Merzenich, M. M. (1998). Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *J. Neurosci.*, *18*(4), 1559–1570.
Ortiz, J. A., & Wright, B. A. (2009). Contributions of procedure and stimulus learning to early, rapid perceptual improvements. *J. Exp. Psychol. Hum. Percept. Perform*, *35*(1), 188–194.
Petersen, S. E., van Mier, H., Fiez, J. A., & Raichle, M. E. (1998). The effects of practice on the functional anatomy of task performance. *Proc. Natl. Acad. Sci. USA*, *95*(3), 853–860.
Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, *256*(5059), 1018–1021.
Pourtois, G., Rauss, K. S., Vuilleumier, P., & Schwartz, S. (2008). Effects of perceptual learning on primary visual cortex activity in humans. *Vision Res.*, *48*(1), 55–62.
Recanzone, G. H., Merzenich, M. M., Jenkins, W. M., Grajski, K. A., & Dinse, H. R. (1992). Topographic reorganization of the hand representation in cortical area 3b owl monkeys trained in a frequency-discrimination task. *J. Neurophysiol.*, *67*(5), 1031–1056.
Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.*, *13*(1), 87–103.
Robinson, K., & Summerfield, A. Q. (1996). Adult auditory learning and training. *Ear Hear.*, *17*(3, Suppl.), 51S–65S.
Roth, D. A., Amir, O., Alaluf, L., Buchsenspanner, S., & Kishon-Rabin, L. (2003). The effect of training on frequency discrimination: Generalization to untrained frequencies and to the untrained ear. *J. Basic Clin. Physiol. Pharmacol.*, *14*(2), 137–150.
Rubin, N., Nakayama, K., & Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Curr. Biol.*, *7*(7), 461–467.
Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nat. Neurosci.*, *8*(12), 1647–1650.
Schiltz, C., Bodart, J. M., Michel, C., & Crommelinck, M. (2001). A pet study of human skill learning: Changes in brain activity related to learning an orientation discrimination task. *Cortex*, *37*(2), 243–265.
Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, *412*(6846), 549–553.
Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Curr. Opin. Neurobiol.*, *17*(2), 148–153.

Seitz, A., & Watanabe, T. (2005). A unified model for perceptual learning. *Trends Cogn. Sci.*, *9*(7), 329–334.

Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron*, *46*(5), 823–835.

Tucker, A., Williams, P. I., & Jeffress, L. A. (1968). Effect of sinal duration on detection for gated and for continuous noise. *J. Acoust. Soc. Am.*, *44*(3), 813–816.

Vaina, L. M., Belliveau, J. W., des Roziers, E. B., & Zeffiro, T. A. (1998). Neural systems underlying learning and representation of global motion. *Proc. Natl. Acad. Sci. USA*, *95*(21), 12657–12662.

van Wassenhove, V., & Nagarajan, S. S. (2007). Auditory cortical plasticity in learning to discriminate modulation rate. *J. Neurosci.*, *27*(10), 2663–2672.

Virsu, V., Oksanen-Hennah, H., Vedenpaa, A., Jaatinen, P., & Lahti-Nuuttila, P. (2008). Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. *Exp. Brain Res.*, *186*(4), 525–537.

Watson, C. S. (1980). Time course of auditory perceptual learning. *Ann. Otol. Rhinol. Laryngol. Suppl.*, *89*, 96–102.

Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *J. Neurosci.*, *17*(10), 3956–3963.

Wright, B. A., & Fitzgerald, M. B. (2001). Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proc. Natl. Acad. Sci. USA*, *98*(21), 12307–12312.

Wright, B. A., & Fitzgerald, M. B. (2005). Learning and generalization of five auditory discrimination tasks as assessed by threshold changes. In D. Pressnitzer, A. de Cheveigne, S. McAdams, & L. Collet (Eds.), *Auditory signal processing: Physiology, psychoacoustics, and models*. New York: Springer.

Wright, B. A., & Sabin, A. T. (2007). Perceptual learning: How much daily training is enough? *Exp. Brain Res.*, *180*(4), 727–736.

Wright, B. A., & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *364*, 301–311.

Yao, H., & Dan, Y. (2005). Synaptic learning rules, cortical circuits, and visual function. *Neuroscientist*, *11*(3), 206–216.

Yao, H., Shi, L., Han, F., Gao, H., & Dan, Y. (2007). Rapid learning in cortical coding of visual scenes. *Nat. Neurosci.*, *10*(6), 772–778.

Yotsumoto, Y., Watanabe, T., & Sasaki, Y. (2008). Different dynamics of performance and brain activation in the time course of perceptual learning. *Neuron*, *57*(6), 827–833.

Zhang, Y., & Wright, B. A. (2007). Similar patterns of learning and performance variability for human discrimination of interaural time differences at high and low frequencies. *J. Acoust. Soc. Am.*, *121*(4), 2207–2216.

Zhang, Y., & Wright, B. A. (in review). An influence of amplitude modulation on interaural level difference processing suggested by learning patterns of human adults.

Zwislocki, J., Maire, F., Feldman, A. S., & Rubin, H. (1958). On the effect of practice and motivation on the threshold of audibility. *J. Acoust. Soc. Am.*, *30*(4), 254–262.

25 Auditory Object Analysis

TIMOTHY D. GRIFFITHS, SUKHBINDER KUMAR, KATHARINA VON KRIEGSTEIN, TOBIAS OVERATH, KLAAS E. STEPHAN, AND KARL J. FRISTON

ABSTRACT The question addressed in this chapter is how the auditory system allows us to represent the elements of the acoustic world? The term *auditory object* is widely used in the literature but in a number of different ways. We consider different aspects of object analysis and the ways in which these can be approached by using experimental techniques such as functional imaging. Functional imaging allows us to map networks for the abstraction of perceived objects and generalization across objects. This fundamental aspect of auditory perception involves high-level cortical mechanisms in the lateral temporal lobe. Systems identification techniques based on Bayesian model selection in individual subjects allow the testing of specific models that explain the activity of the networks that are mapped.

The concept of auditory object

In the acoustic world, we experience a number of different things that form the natural sound scene. The problem considered here is how the brain abstracts representations of these things, or objects, as a basis for perception. The computation required for this process is formidable, given the richness of our sound experience that is entirely based on two pressure waveforms arriving at the ears. The problem is a key issue for what has become known as auditory scene analysis (Bregman, 1990).

In contrast to the concept of visual objects, the concept of auditory object is controversial for a number of reasons (Griffiths & Warren, 2004). At the level of the stimulus, it is more difficult to examine the sound pressure waveform that enters the cochlea and "see" different objects in the same way that we "see" objects in the visual input to the retina. However, in the auditory system and in the visual system, objects can be understood in terms of the images they produce during the processing of sense data. The idea that objects are mental events that result from the creation of images from sense data goes back to Kant and Berkeley (Russell, 1945). In the visual system, there is good evidence for the creation of images (brain representations corresponding to an object) with two or more spatial dimensions in the form of arrays of neural activity that preserve spatial relationships from the retina to the cortex. In the auditory system, the concept of an image is most often used to refer to a brain representation with dimensions of frequency and time or derivations of these such as spectral ripple density related to frequency (Chi, Ru, & Shamma, 2005) and amplitude modulation (Chi et al., 2005) or forms of autocorrelation (Patterson, 2000) related to time. If we accept the existence of images with a temporal dimension, then the concepts of auditory objects and auditory images can be considered in a way comparable to how the visual system is considered. The idea was first proposed by Kubovy and Van Valkenburg (2001), who suggested that auditory objects can be considered as existence regions within frequency-time space that have borders with the rest of the sound scene.

A second issue about the concept of auditory object analysis (which is also relevant to visual object analysis) is the cognitive level to which it should be extended. Consider the situation in which you hear someone making the vowel sound /a/ at a pitch of 110 Hz and intensity of 75 dB on the left side of the room. That situation requires sensory analysis of the spectrotemporal structure of the sound. It also requires categorical perception to allow the sound to be distinguished from other sounds. Sounds from which it has to be distinguished might be from another class (e.g., a telephone ringing at the same pitch, intensity, and location) or the same class (e.g., another person making the vowel sound /a/ at a different pitch, intensity, or spatial location). We can appreciate that we are listening to the same type of sound if we hear it at 80 Hz or 65 dB or on the right side of the room. We can appreciate that similarity even if we do not speak a relevant language to allow us to recognize or name the vowel. At another level of analysis, the sound must enter a form of echoic memory store (to allow comparison with sounds that might immediately follow it) and might enter an anterograde memory store that allows comparison with sounds heard over days or weeks. At a further level of analysis, we might call the sound a voice, or my voice, the vowel "a," or (if we have absolute pitch) "A2." The term *object analysis* might therefore be applied to (1) the perception of a coherent whole, the essence of

TIMOTHY D. GRIFFITHS and SUKHBINDER KUMAR Institute of Neuroscience, Newcastle University, Newcastle upon Tyne; Wellcome Centre for Imaging Neuroscience, University College, London, United Kingdom

KATHARINA VON KRIEGSTEIN, TOBIAS OVERATH, KLAAS E. STEPHAN, and KARL J. FRISTON Wellcome Centre for Imaging Neuroscience, University College, London, United Kingdom

which can be perceived even when cues such as pitch or intensity are changed; (2) categorical analysis; (3) encoding into working memory; (4) encoding into anterograde memory; or (5) association with a label during semantic analysis. We prefer to apply the term *object analysis* to the first stage and will emphasize it in the current chapter, but a number of workers would argue for an obligatory requirement for auditory objects, or objects in general, to have an associated label. The key point, however, is that a number of stages of analysis are required to assess the nature of an auditory object and that even the first stage above requires considerable computational work to derive a representation of particular sounds that is independent of basic cues such as pitch or intensity. Clinically, the distinction of presemantic and semantic processing stages of object analysis is relevant to the existence of apperceptive and associative forms of auditory agnosia, respectively (Griffiths, Bamiou, & Warren, in press).

A third controversial aspect of auditory object analysis is whether the concept should be applied to particular individual sounds that can be distinguished from others as argued above or to *sequences* of sounds that are grouped: auditory streams. Bregman (1990) explicitly rejects the concept of auditory object in favor of the auditory stream, a sequence of grouped sounds, as the fundamental unit of auditory perception. Others equate auditory objects with streams (Shamma, 2008). There is a problem, however, with considering streams as auditory objects corresponding to a single percept derived from analysis over longer periods of time. Streams are sequences of sounds that are grouped by perceptual properties such as pitch, timbre, or position (Moore & Gockel, 2002), where these perceptual properties all have complex relationships to the acoustic structure. This makes a description of the stream as the most fundamental unit of analysis leading to perception problematic, when the stream itself comprises elements that are perceived individually. Whether streams are regarded as objects, streams of objects, or something else, however, it is an important level of perceptual organization that will also be considered here.

In this chapter we will consider different ways in which auditory object analysis can be approached, with an emphasis on human neuroimaging. We will develop an approach that allows an understanding of how we assess objects as perceptual wholes that can be distinguished from others and generalized. We will also consider the encoding and retrieval of sequences of objects. Functional neuroimaging data demonstrate that such auditory object analysis requires distributed networks in the temporal lobe distinct from early mechanisms for the representation of spectrotemporal stimulus structure and perceived pitch in the superior temporal plane. We now have tools to tease out the detailed functional organization of the networks for object analysis.

Natural stimuli for the investigation of auditory object analysis

A reasonable starting point for the analysis of auditory objects is the use of sampled natural stimuli. These have been used in behavioral experiments that use the technique of multidimensional scaling (MDS) (Caclin, McAdams, Smith, & Winsberg, 2005; Grey, 1977). This involves pairwise judgments of the degree of difference between sounds that are then arranged in an *n*-dimensional space, where the degree of dissimilarity between sounds is represented by the degree of physical separation. The best fit for a given number of dimensions is determined by minimizing the normalized sum of squares error for the fits that can be expressed as a stress factor. Figure 25.1 shows musical instruments placed in a three-dimensional Euclidian space using the technique. The technique is not dependent on any assumptions about what causes them to sound different. However, the dimensions disclosed by MDS can be examined to determine whether the dimension corresponds to any systematic variation of acoustic properties. For musical sounds with the same pitch, the dimensions of timbre space have been argued to correspond to the spectrum of the sound (characterized by the spectral centroid), the temporal envelope (attack and decay), and changes in spectral centroid over time. Recent work (Caclin et al., 2005) has also emphasized the relevance of the fine-spectral structure.

Synthetic manipulation of natural stimuli to investigate auditory object analysis

An alternative and complementary approach to sampling natural sounds is to alter natural sound objects systematically using algorithms that obey natural principles. Figure 25.2 shows how this approach can be applied to sounds produced by natural resonant sources (a human voice, a French horn, and a bullfrog). Natural resonant sources usually produce a fundamental frequency and a series of harmonics at integer multiples of the harmonic frequency; the sound has a number of components when considered in the spectral domain. These components are filtered by the resonant cavities within the source, which has the effect of producing peaks or formants in the spectrum. Figure 25.2 shows modeled auditory images (Griffiths, Buchel, Frackowiak, & Patterson, 1998) for the three types of sound, where the activity within the fibers of the auditory nerve (arranged in increasing frequency order on the vertical axis), undergoes a form of autocorrelation (shown on the horizontal axis), as might be used in the auditory system to achieve time stabilisation of the auditory image. The vertical ridge at a time delay of 6.8 ms corresponds to the temporal regularity within these stimuli, which all have the same pitch (147 Hz). The formant peaks can be seen as horizontal bands of

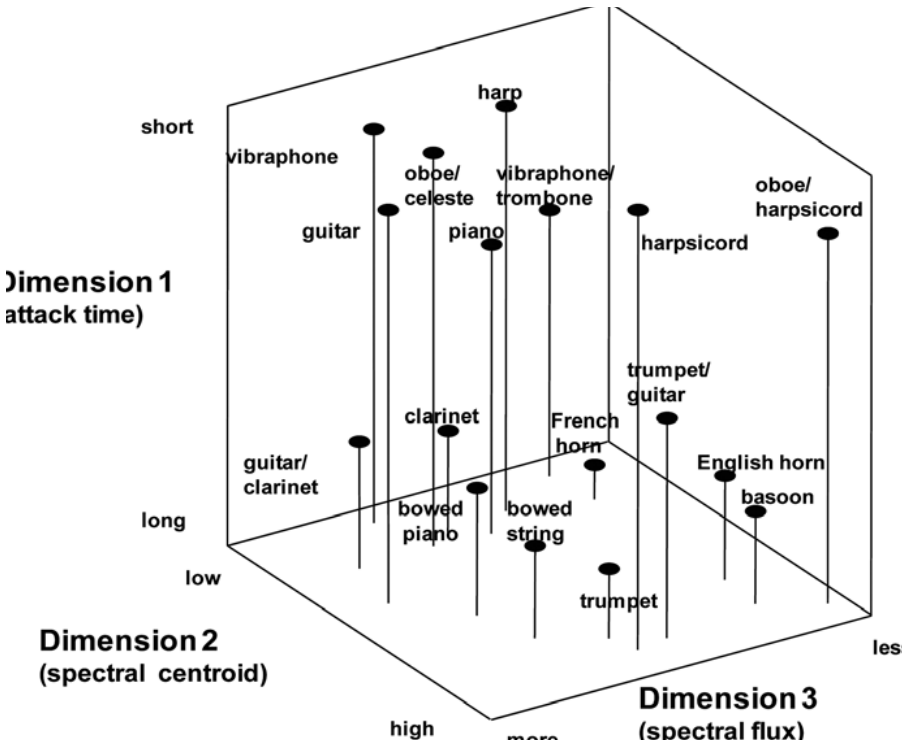


FIGURE 25.1 Multidimensional scaling to disclose the perceptual relationships between natural stimuli (Menon et al., 2002).

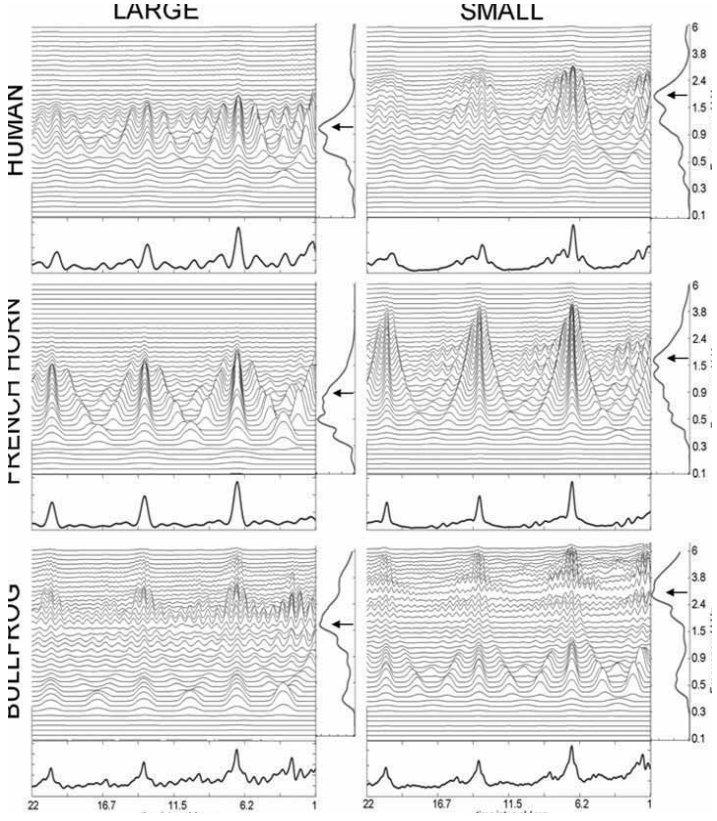


FIGURE 25.2 Resonator scale stimuli (von Kriegstein et al., 2007).

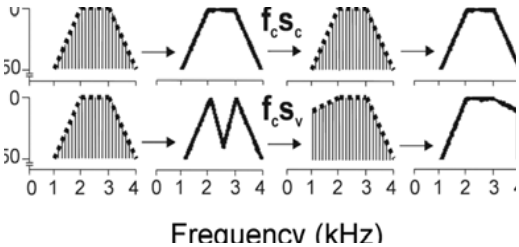


FIGURE 25.3 Spectral envelope stimuli (Warren et al., 2005).

activity at particular frequencies that are seen at the right of each figure in the form of a mean spectral activity pattern. The two columns show the auditory image for sounds emitted by large sources on the left and small sources on the right. When the sound source gets smaller, the formant peaks move to a higher frequency, as shown by the position of the black arrow, corresponding to one of the formants. Notice that this can happen without changing the pitch. Behavioral experiments (Smith, Patterson, Turner, Kawahara, & Irino, 2005) demonstrate that in the case of voices, humans judge the size of auditory objects using the acoustic effect of resonator size to a greater extent than changes in pitch.

We have used an algorithm called STRAIGHT (Kawahara & Irino, 2004) to modify the acoustic correlates of resonator size (figure 25.2). Perceptually this produces a different type of object change from the instrument changes in figure 25.1, where the perceived object remains in the same class. Unlike pitch change, change of resonator size does not generally occur within one source (with some interesting exceptions, such as stags that drop their larynges when calling during the rutting season (Fitch & Reby, 2001)). Size changes are generally perceived as object changes when they can be detected, which makes sense ethologically. Other experimental manipulations of natural objects that have been used to assess object analysis experimentally include morphing or mixing techniques similar to those used in vision (e.g., Zatorre, Bouffard, & Belin, 2004).

Synthetic stimuli for the investigation of auditory object analysis

Work based on MDS has identified a key spectral dimension that determines the differences between objects. Figure 25.3 shows synthetic stimuli that have been used in imaging experiments to reveal mechanisms for the analysis of perceived changes in objects related to the spectral envelope, regardless of changes in the fine spectral structure. The work is an example of an approach that might be called *prototimbre* based on the systematic manipulation of timbral dimensions (see Griffiths, 2008, for further examples). The work seeks mechanisms comparable to those in vision that allow the same face to be perceived regardless of angle or illumination. The use of synthesis allows the systematic manipulation of dimensions identified by MDS in a more straightforward way than using sampled stimuli. Consider the top row of stimuli in figure 25.3. This shows the spectral representation of successive sounds in which the fine structure alternates between harmonic sounds (with an associated pitch) and noise. A common spectral envelope is applied to the sounds and listening to the stimuli establishes that there is an essential "sameness" despite the very different spectral structure. A similar mechanism allows us to perceive the same vowel sound whether it is voiced or whispered. The sounds in figure 25.3 could be vowels but not any that you might have heard, and the use of stimuli at this level allows an assessment of mechanisms relevant to generalization across different fine structure without any semantic association. The lower row of figure 25.3 shows the situation in which the spectral envelope changes from one sound to the next. This is perceived as a changing object over and above the changing fine structure. Measurements of brain activity in response to the lower stimulus compared with the upper stimulus allows inference about mechanisms for the abstraction of object identity over and above the analysis of the fine spectrotemporal structure.

Stimuli based on sequences of objects

A number of behavioral studies have examined sequential grouping based on pitch since the original demonstration of this in a paradigm introduced by van Noorden (1975). The studies demonstrate that in a sequence of the form A-B-A-A-B-A, where A and B are different pitch values, sequential grouping will occur (we hear "horse" based on sequentially grouped A-B-A triplets rather than "morse" based on segregated A and B streams) when the pitch separation between the high and the low pitch is not large and the rate of presentation of individual notes is not high (Bregman, 1990). Sequential grouping of sounds (the opposite of which, segregation, is called *streaming*) can be achieved on the basis of a number of perceptual properties, including pitch, in the majority of experiments in addition to timbre and position in space (Moore & Gockel, 2002).

In the natural world, we do not hear deterministic sequences; we hear stochastic sequences with varying degrees of predictability. Figure 25.4 shows synthetic pitch sequences based on pitch trajectories that are derived from power spectra that are related to the frequency, f, by the function f^{-n}. For any given value of n, families of pitch sequences with similar statistical properties can be constructed on the basis of power spectra with that value of n: different exemplars are created by the use of different random-phase spectra. When $n = 0$ (top part of figure 25.4), the power spectrum is flat, and the pitch trajectory created corresponds to fixed-

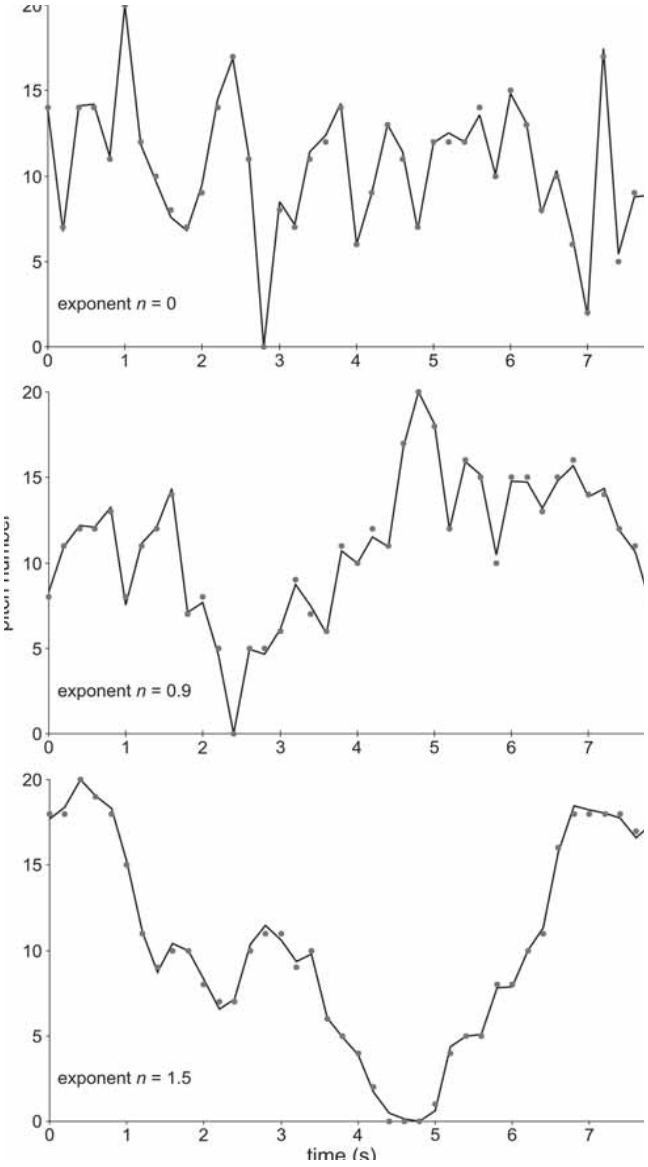


FIGURE 25.4 Fractal pitch stimuli with varying information content used to probe for sequence encoding mechanisms (see text). The pitch trajectories were derived from power spectra with the form f^{-n}, where the value of the exponent n is given with each figure. Pitch number refers to a pitch scale spanning two octaves, where each octave is divided into 10 equal log divisions. Low exponents (top) produce unpredictable pitch sequences with a large amount of information, while high exponents produce redundant sequences containing less information.

amplitude random-phase noise. In such a waveform, successive pitch values cannot be predicted by the preceding pitch, so each pitch value contains a lot of information. In contrast, as n tends to a large value, the pitch waveform tends to a sine wave in which successive pitch values can be accurately predicted by the preceding pitch, and individual pitch values (or sequences of a given length) do not contain a large amount of information. These stimuli are called *fractal pitch sequences* because of their scaling properties (Schmuckler & Gilden, 1993), but the important property used in our experiments is the relationship between n and the information contained in a sequence. Specifically, we hypothesized that computationally efficient mechanisms for the encoding of sequences of sound would use less computational resource as the amount of information in sequence decreased. These sounds are ethological in that they contain a global contour, on which smaller excursions are imposed as occurs in a variety of natural stimuli. Sampling of natural acoustic patterns demonstrates a similar balance: Music and speech have trajectories for pitch and other perceptual properties corresponding to f^{-n} power spectra, where $n = 1$ (Voss & Clarke, 1975).

Strategies for the brain measurement of auditory object analysis

A number of techniques can be used to assess brain processes related to auditory object analysis, from single-unit approaches to measurements reflecting ensemble electrical activity in MEG and EEG experiments and approaches based on the BOLD response in fMRI (considered in the next section), which reflects blood flow changes in response to neuronal ensemble activity as indexed by local field potentials (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). Similar principles might be applied to all of the techniques to allow inference about object analysis. At the level of single units, a helpful approach has been to characterise the selectivity of units to particular types of natural objects by establishing a profile of responses to different exemplars (Tian, Reser, Durham, Kustov, & Rauschecker, 2001; Kikuchi, Horwitz, & Mishkin, 2007). The approach has suggested a gradient of object selectivity within the superior temporal lobe with more selective responses toward the anterior temporal pole. At the level of ensemble activity, responses to natural classes of stimuli might also be sought as in the case of imaging studies of voices considered in the next section, and there are now also fMRI techniques available to establish the mapping of different exemplars within the same human cortical areas (see below). In the case of both single-unit and ensemble activity, a critical issue for natural stimuli is whether the responses are true object responses that reflect abstraction of object properties rather than a simple representation of particular spectrotemporal features. Another critical issue for natural stimuli is whether the representation is at the level of the "coherent whole" that is perceived or at the level of the associated semantic label. Definition of responses that are specific to object analysis could be achieved with any technique by defining responses that occur for between-object acoustic change (such as /a/ to /e/) but not for within-object acoustic change (such as /a/ at a different level or pitch). That approach utilizes the phenomenon of object constancy to define object-specific mechanisms.

Univariate analysis of fMRI data: Single objects

Functional imaging that assesses the BOLD response during object analysis using mass-voxel-wise univariate statistics has used a number of different stimuli, including natural sounds, synthetically manipulated natural sounds, and synthetic sounds. Figure 25.5 shows an example of an experiment in which a response to a category of natural sound was sought (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). In this experiment, the BOLD response during the perception of voices (both speech and nonspeech sounds) was contrasted with responses to nonvocal environmental sounds and demonstrated bilateral responses in the superior part of the superior temporal sulcus (STS). Figure 25.6 shows a mapping of changes in the BOLD response when subjects listen to variation in the size and type of resonator using stimuli shown in figure 25.2. During this experiment, subjects perceived successions of different-sized voices, French horns, or bullfrogs (von Kriegstein, Smith, Patterson, Ives, & Griffiths, 2007). Perception of object change within categories produced in this way was associated with bilateral increases in the BOLD signal in the superior temporal gyrus: this experiment also provided evidence for more specific change in the left posterior superior temporal gyrus for changes in the size of voices. The experiment demonstrated that very similar regions respond to changes in category as to changes in resonator size. The data can be interpreted in terms of generic mechanisms for the analysis of object change, whether this corresponds to a change in resonator size or a change in the class of sound.

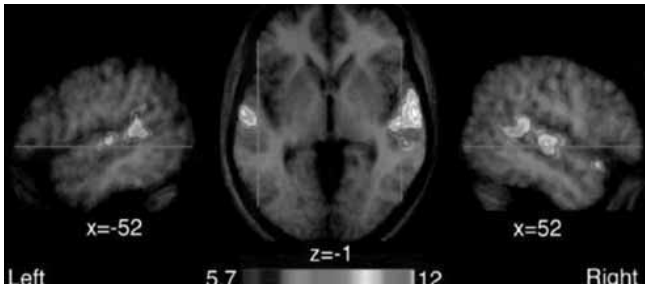


FIGURE 25.5 Activation during passive listening to voices (Belin et al., 2000). The numbers refer to planes defined in millimeters in Talairach space (Talairach & Tournoux, 1988). (See color plate 28.)

In the resonator-size changes or category changes (von Kriegstein et al., 2007), there are associated changes in the spectral and temporal structure of the stimulus. Figure 25.7 shows an experiment in which the basis for the analysis of spectral envelope in generic acoustic objects (without associated meaning) is assessed (Warren, Jennings, & Griffiths, 2005). The figure shows mapping within the superior tem-

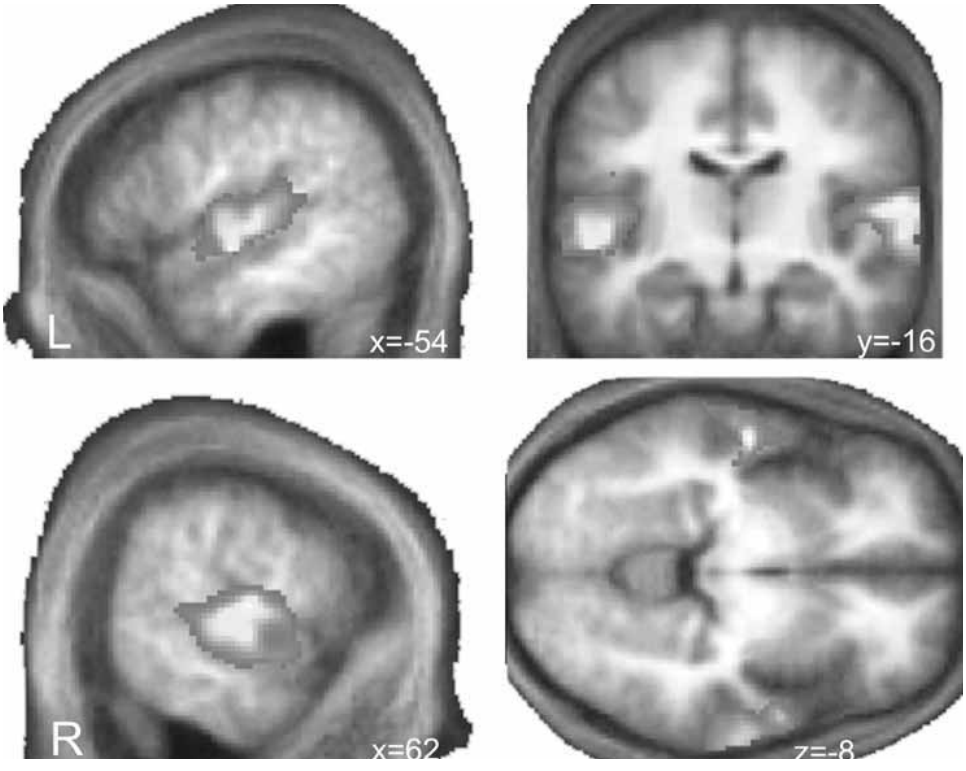


FIGURE 25.6 Activation due to passive listening to changing resonator scale and sound class in the three types of harmonic sounds shown in figure 25.2 (von Kriegstein et al., 2007). (See color plate 29.)

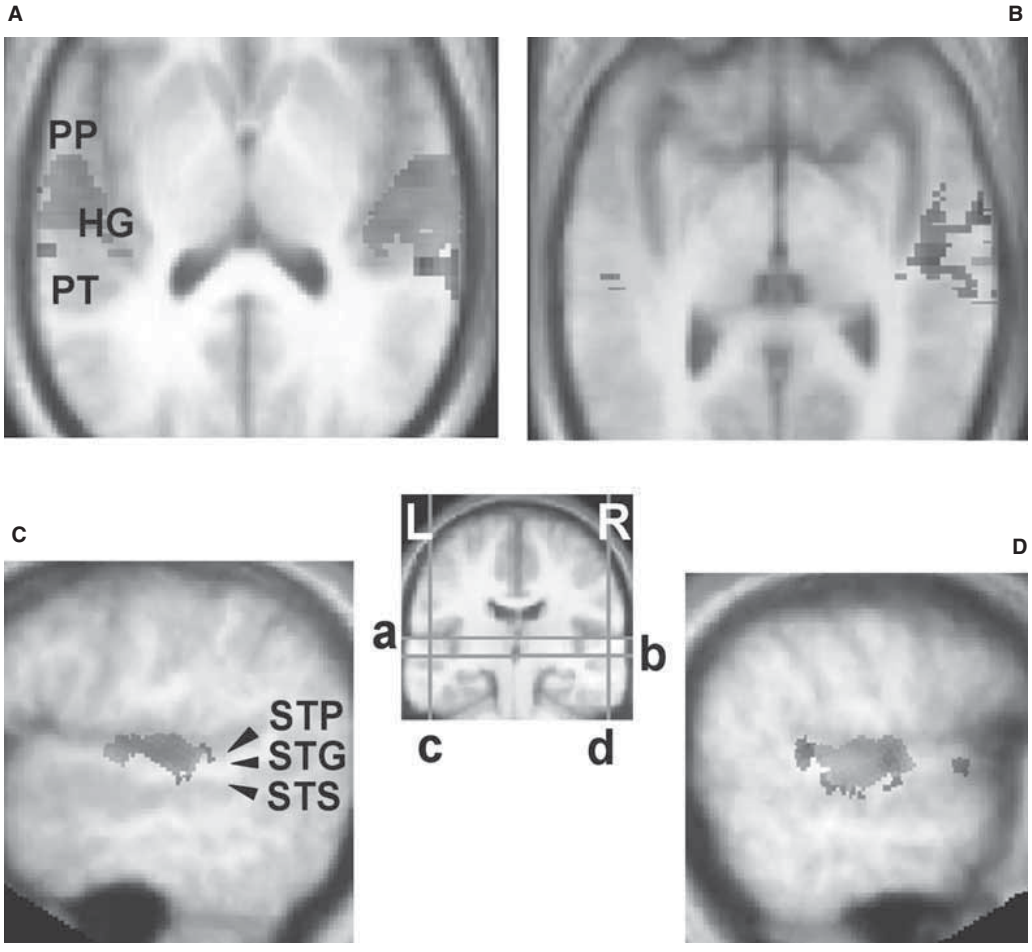


FIGURE 25.7 Activation due to passive listening to changing spectral envelope (Warren et al., 2005). HG, Heschl's gyrus; PT, planum temporale; PP, planum polare; STS, superior temporal sulcus. (See color plate 30.)

poral plane (in red) corresponding to whether or not the sounds were associated with pitch. That mapping occurs in lateral Heschl's gyrus (HG) in a region previously demonstrated to increase activity as a function of pitch salience (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Penagos, Melcher, & Oxenham, 2004). The key contrast in figure 25.7 (in blue) is between changing spectral envelope and fixed spectral envelope in series of objects with continuously varying fine-spectral structure (shown in figure 25.3): an argument can be made that this contrast identifies areas involved in the "abstraction" of spectral envelope relevant to object analysis over and above the analysis of the fine-spectral structure. The contrast shows bilateral activation in the superior temporal plane in the planum temporale (PT), posterior to the pitch mechanisms, and predominantly right-lateralised activation in the STS.

These studies all highlight a critical role in object analysis for temporal lobe areas beyond the primary and secondary cortices in HG in the superior temporal plane. The areas are likely to be involved in the abstraction of object properties beyond the representation of spectrotemporal structure. We consider later how the responsible system for spectral envelope analysis might be determined explicitly by using dynamic causal models of functional auditory architectures.

Univariate analysis of fMRI data: Sequences of objects

Figure 25.8 shows an experiment in which the encoding of sequences of objects was assessed: specifically, the encoding of the fractal-pitch sequences similar to the examples in figure 25.4. The information content of a pitch series was systematically varied by changing the exponent, n, determining a power spectrum with the form f^{-n} from which the pitch series was derived. The experiment was carried out as an explicit search for mechanisms for the encoding of auditory sequences. It was predicted that computationally efficient encoding mechanisms should use less computational resource (measured indirectly by using the BOLD response) for more

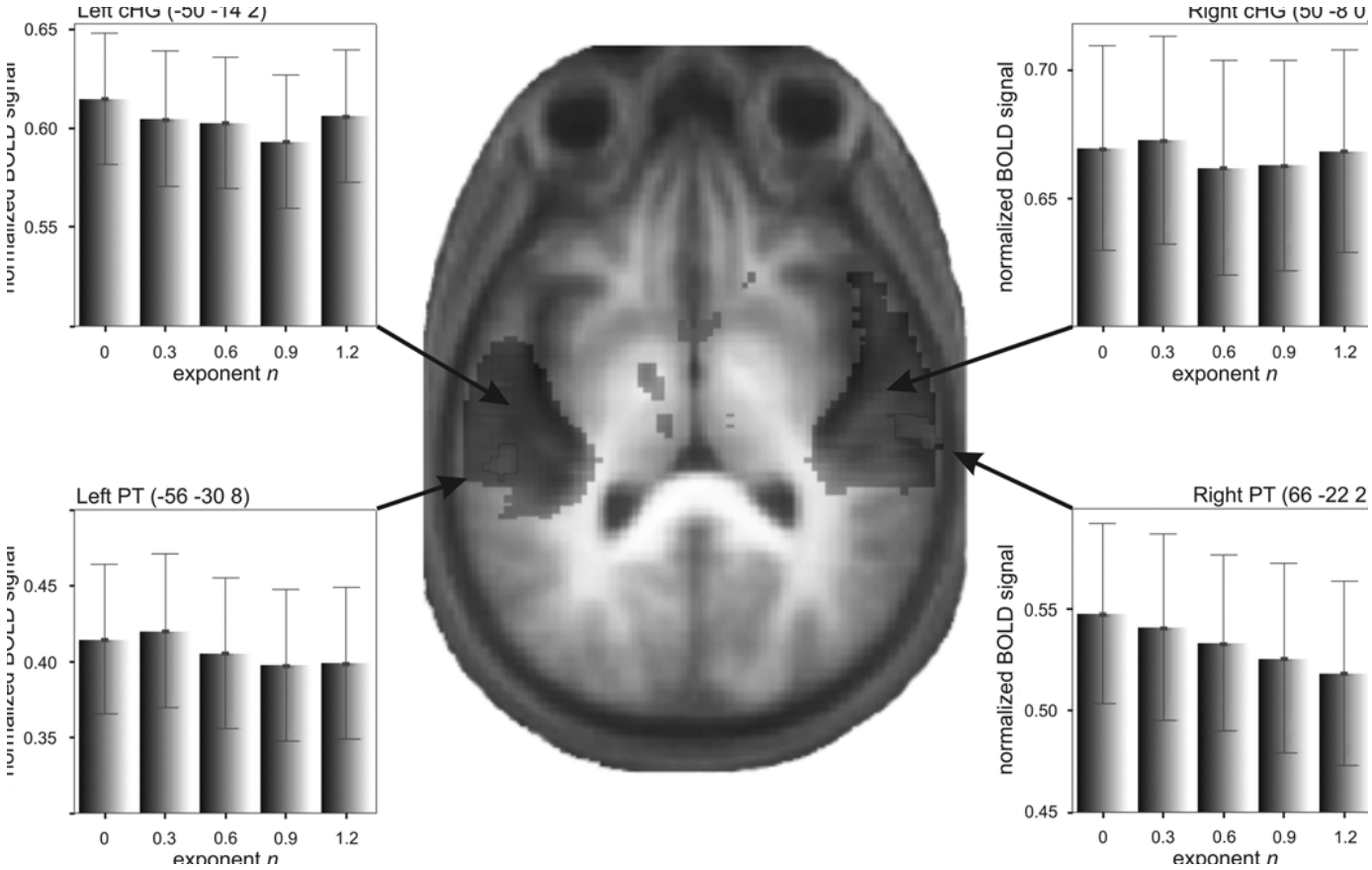


FIGURE 25.8 Activation as a function of information content of pitch sequences during the encoding of pitch sequences. A significant effect of pitch-sequence information content (which decreases as *n* increases) is shown in the planum temporale (PT) but not in Heschl's gyrus (HG). Numbers in parentheses are Talairach coordinates in millimeters where the BOLD values were measured. (See color plate 31.)

redundant sequences containing less information. Such a relationship was demonstrated in two experiments in the PT, bilaterally, but not in the primary and secondary auditory cortices in HG. The work is consistent with the suggestion (Griffiths & Warren, 2002) that the PT represents a "computational hub" responsible for the encoding of acoustic stimuli, and suggests overlapping substrates for the abstraction of object features as in figure 25.7 and the encoding of sequences of objects. In contradistinction, figure 25.9 shows a contrast to demonstrate areas involved in the retrieval of auditory sequences in the second experiment during a one-back task where subjects were required to compare successive pitch sequences. The contrast demonstrates bilateral frontal activity including activity in the frontal operculum which in the right hemisphere is similar to that occurring during working memory tasks for melodic pitch sequences (Zatorre, Evans, & Meyer, 1994). Unlike encoding, the activity associated with retrieval was not affected by the information content of the stimulus. This can be interpreted in terms of the retrieval process requiring a symbolic level of processing that is not yoked to the complexity of the acoustic stimulus in the same way as encoding.

Multivariate analysis of fMRI data

There has been considerable interest in techniques to demonstrate different spatial distributions of BOLD activity in response to sensory stimulation, which can be achieved by the use of multivariate statistical methods. For a description of this approach to visual data, see Haynes and Rees (2006). The technique has the potential resolution to allow fMRI characterization of different responses within the same cortical areas that correspond to the perception of different individual auditory objects. The interpretation of such mappings would be subject to the same issues discussed above in terms of whether spectrotemporal structure or a correlate of the perceived object is represented.

Analysis of categorical processing using fMRI

Categorical response to changes in objects can be assessed by using the technique of repetition suppression that has been developed for the analysis of visual fMRI data (Grill-Spector, Henson, & Martin, 2006). Previous work suggested categorical mechanisms for visual representation based on

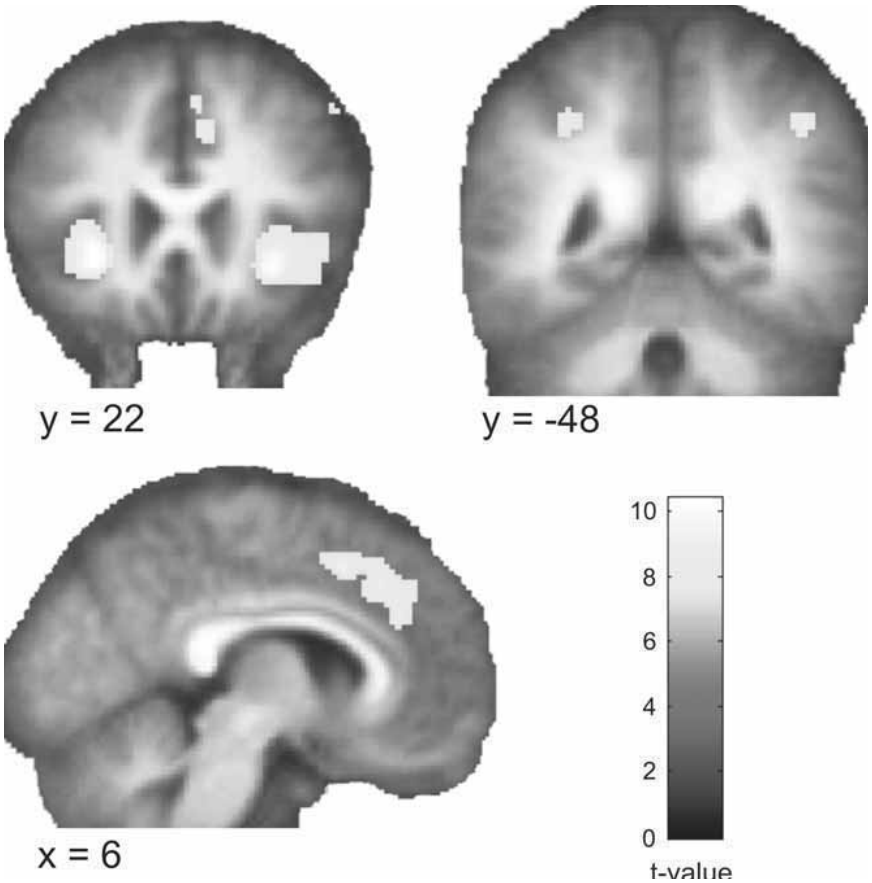


FIGURE 25.9 Contrast to demonstrate activation associated with pitch-sequence retrieval in an active listening task. (See color plate 32.)

BOLD responses to exemplars from the same category that decrease with repeated presentation, regardless of other (category-independent) stimulus changes. The technique allows categorical mechanisms to be sought even when different neuronal ensembles tuned to different categories are located in the same region. Recent visual neurophysiological work (Sawamura, Orban, & Vogels, 2006) demonstrates correlates of the phenomenon at the single-unit level. Models that might explain the phenomenon at the neuronal ensemble level are developed in Grill-Spector et al., (2006). A recent study applied a related approach to mapping of an auditory continuum between two phonemes (Raizada & Poldrack, 2007). That study demonstrated responses that changed across phoneme boundaries in areas beyond the temporal lobe, but the technique could also be applied to shifts between objects at a presemantic level that might be analyzed in temporal lobe areas.

Effective connectivity analysis of fMRI data

The conventional analyses considered in figures 25.5 to 25.9 demonstrate considerable overlap in the networks of activity that are involved in the analysis of objects assessed using different types of stimulus manipulation and in the analysis of sequences of objects. In particular, a key role for the PT is demonstrated in these studies consistent with the idea that this is an important "computational hub" concerned with auditory encoding. The term *hub* implies connection to other nodes of analysis and a flow of information: There is a need for the identification of specific systems for object analysis that might use similar nodes in different ways. Specifically, different aspects of object analysis might be subserved by different patterns of connectivity between nodes. In this section, we consider the application of this approach to one aspect of object analysis, spectral envelope analysis, addressing the question of how PT and the other nodes within the right-hemisphere network for spectral envelope analysis are effectively connected.

We use an approach called dynamic causal modeling (DCM) (Friston, Harrison, & Penny, 2003), together with Bayesian model selection (Penny, Stephan, Mechelli, & Friston, 2004), to test different models for auditory object analysis. The approach identifies effective connectivity between areas (the causal influence of activity in one area on the activity in another) and the modulatory effect of task (or any other experimentally controlled manipulation) on effective connectivity. DCM belongs to a family of models of effective connectivity such as structural equation

modeling (SEM) (McIntosh & Gonzalez-Lima, 1994), multivariate autoregression (Harrison, Penny, & Friston, 2003), or Granger causality (Goebel, Roebroeck, Kim, & Formisano, 2003). DCM has a number of advantages (discussed in Friston et al., 2003; Stephan, 2004). A major advantage is the way in which DCM can be used to carry out a systematic comparison of competing models that might explain a regional pattern of BOLD activity.

The basic idea behind dynamic causal modeling can be summarized as follows. A cognitive or motor task in the brain is accomplished by interaction between a number of nodes. This interaction is at the level of neural activity and therefore takes place at the millisecond time scale. DCM models these neuronal interactions. However, except for invasive recording studies (e.g., Moran et al., 2008), we cannot directly observe neuronal activity, but only some consequence of it, such as a hemodynamic BOLD signal or an EEG signal measured at scalp sensors. For this reason, DCM combines a model of neuronal dynamics with a biophysical forward model that explains how the hidden neuronal activity translates into a measured signal (Kiebel, David, & Friston, 2006; Friston, Mechelli, Turner, & Price, 2000; Stephan, Weiskopf, Drysdale, Robinson, & Friston, 2007). The incorporation of the biophysical model allows inferences to be made from coarsely sampled BOLD time series about neuronal events occurring at a much finer time scale. With respect to auditory functional MRI, many experiments, including our experiment on spectral envelope analysis above, are based on "sparse" designs to avoid the effect of simultaneous scanner noise on the effects of interest. Despite the sampling rate for "sparse" BOLD time series approaching 0.1 Hz, plausible models of dynamic neural interactions at the millisecond level can still be disambiguated, given the fMRI data. This is because the forward model predicts, given the known experimental inputs, what the BOLD signal should look like at any future time point, including the times when BOLD measurements were taken; the sampling frequency (repetition time) is irrelevant.

Like any model, DCM comprises variables (that may or may not be measurable) and parameters that are estimated from the measurements. The model that is used in DCM has three types of variables: input variables (the same as those used in conventional analyses based on the general linear model, or GLM), encoding the experimental manipulation; output variables that are the regional hemodynamic responses from each of the regions considered in the model; and state variables. State variables describe the "hidden" (unobserved) states of the system and represent the neural activity and biophysical variables (e.g., blood flow) that transform neural activity into a hemodynamic response. DCM uses three different sets of parameters: endogenous parameters that model the baseline connection strengths between the regions in the absence of any external excitation of the system, modulatory parameters that model the change in endogenous connection strength induced by the external experimental input, and a third set of parameters that model the direct influence of an exogenous stimulus on a given region. The conventional GLM analysis is based on the assumption that any exogenous stimulus has a direct influence on a region and therefore it is the third set of parameters that form the primary focus of traditional GLM analyses. DCM, therefore, can also be regarded as a generalization of the GLM in which coupling parameters between regions are allowed to be nonzero.

Once the model has been specified, it has to be estimated from the measurements. There are, however, some natural constraints on the model; for example, the neural activity of a region cannot diverge to infinity. One framework for estimating the parameters with prior constraints is Bayesian statistics, in which a parameter is treated as a random variable that is completely characterized by its probability density (distribution) function. The prior constraints about the parameters are specified in terms of a (prior) density function. Bayesian estimation procedures estimate the parameters in terms of their posterior density function.

For fMRI, DCM is based on a bilinear model of neural population dynamics that is combined with a hemodynamic model (Buxton, Wong, & Frank, 1998; Friston et al., 2000), describing the transformation of neural activity into predicted BOLD responses. The neural dynamics are modeled by the following bilinear differential equation

$$\frac{dz}{dt} = Az + \sum_{j=1}^{m} u_j B^{(j)} z + Cu \qquad (1)$$

where z is the state vector (with one state variable per region), t is continuous time, and u_j is the jth input (i.e., some experimentally controlled manipulation). This state equation represents the strength of connections between the modeled regions (the endogenous A matrix), the modulation of these connections as a function of experimental manipulations (e.g., changes in task; the modulatory or bilinear $B^{(1)} \ldots B^{(m)}$ matrices), and the strengths of direct inputs (e.g., sensory stimuli, the exogenous C matrix). These parameters correspond to the rate constants of the modeled neurophysiological processes. Combining the neural and hemodynamic model into a joint forward model, DCM uses a Bayesian estimation scheme to determine the posterior density of the parameters. Under Gaussian assumptions, this density can be characterized in terms of its maximum a posteriori estimate and its posterior covariance. The parameters of the neural and hemodynamic model are fitted such that the modeled BOLD signals are as similar as possible to the observed BOLD responses. This allows one to understand and make statistical inferences about regional BOLD responses in terms of the connectivity at the underlying neural level.

We used DCM in this way to identify the system for object analysis in the case of spectral envelope analysis shown in figure 25.7 (Kumar, Stephan, Warren, Friston, & Griffiths, 2007). In that experiment, activation during sound perception was seen in the primary and secondary areas in HG and association cortex in the PT. A contrast to show activity corresponding to the perception of object change identified bilateral activity in PT and right-lateralized activity in STS. DCM was used to identify the system that explains the pattern of activity in HG, PT, and the STS in the right temporal lobe. A critical question is whether analysis in PT and STS occurs in a serial fashion, based on connections from HG to PT and from PT to STS, or whether the analysis is based on parallel processing due to connections from HG to both PT and STS. The analysis also addressed how connection strengths between elements of this cortical network are modulated during the spectral envelope analysis.

To test these hypotheses, two broad categories of models, serial and parallel, were specified (figure 25.10). All the models specified were based on the conventional assumption

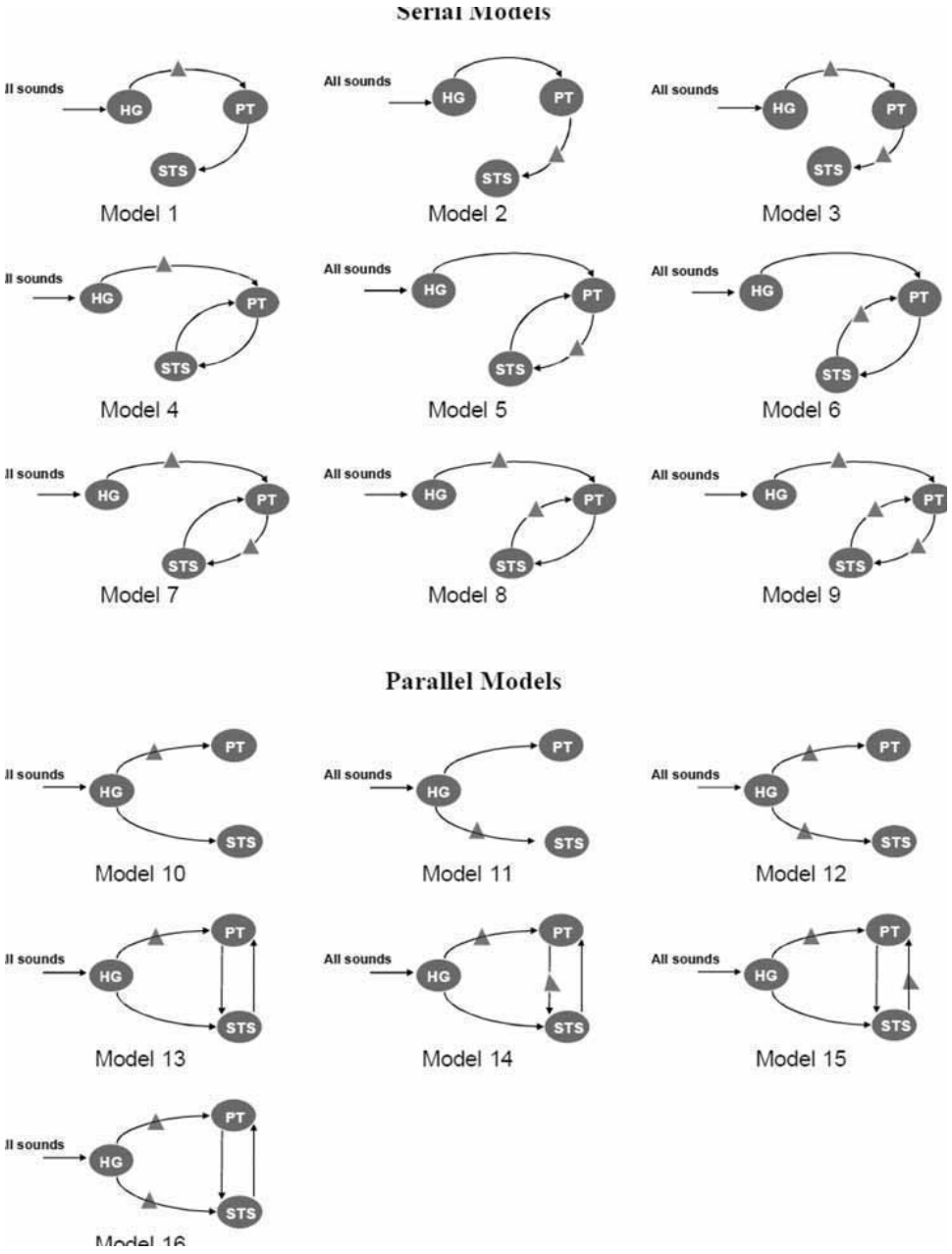


FIGURE 25.10 Serial and parallel models for spectral envelope analysis in the right hemisphere. The triangle in the pathway between two regions indicates the modulatory effect of extraction of spectral envelope (Kumar et al., 2007). HG, Heschl's gyrus; PT, planum temporale; STS, superior temporal sulcus.

that there is a direct or exogenous effect of the sound input on the activity within the primary auditory cortex within HG. In the serial models, auditory inputs entering HG reach STS via PT and thus processing in STS depends on inputs from PT. In contrast, in the parallel models, HG connects to both PT and STS enabling parallel processing in PT and STS. In total, 70 models were fitted to the data (16 of which are shown in figure 25.10) and compared by using Bayesian model selection. It should be noted that the DCM approach developed below might yield a "best" model that is not a "true" model if the set of models tested does not include the latter. It is critical in DCM, therefore, to consider all possible models in a systematic and inclusive way. Even for a simple serial and parallel comparison for three areas, as here, there are a large number of models when all the possible forward and back projections and the possible sites of modulatory effect are taken into account.

A general problem that arises in any modeling exercise is to decide, given a measured data set, which of several competing models is the optimal. A number of criteria for selecting the optimal model have been proposed in the modelling literature (Burnham & Anderson, 2004). The optimal criterion is the model evidence, defined as the probability $p(y|m)$ of obtaining the data y given a particular model m (Raftery, 1995). Critically, the model evidence takes into account not only the relative fit of competing models but also their relative complexity, determined by the number of free parameters. This is important because there is a tradeoff between the fit of a model and how well it might generalize—in other words, how well it explains different data sets generated from the same underlying process. As the number of free parameters is increased, model fit increases, whereas beyond a certain point, the ability of the model to generalize decreases. The reason for this is overfitting: An increasingly complex model will, at some point, start to fit noise that is specific to one data set and thus become less applicable to multiple realizations of the same underlying generative process.

Because the model evidence cannot always be derived analytically, two commonly used approximations to it are Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Penny et al., 2004). These approximations do not necessarily give identical results because BIC favors simpler models, whereas AIC is biased toward more complex models. A general convention is that if two models (say, m_1 and m_2) are to be compared, then a decision is made only when AIC and BIC concur. In either case, the relative evidence of one model as compared to another is determined by the Bayes factor:

$$BF_{12} = \frac{p(y|m_1)}{p(y|m_2)} \qquad (2)$$

where BF_{12} is the Bayes factor of model 1 with respect to 2. Following the selection of a best model for each individual subject, the optimal model for a group of subjects can be determined by the group Bayes factor (GBF), which is equal to the product of the Bayes factors for each individual subject.

Figure 25.11 shows the evidence for the models, determined separately by using AIC and BIC, in a group of eight subjects. Model 1 is the optimal model to explain the data. The parameters for this model specify a serial model with connectivity (HG $\rightarrow$ PT $\rightarrow$ STS) and modulation of connection from HG to PT during the analysis of spectral envelope. Table 25.1 shows the group Bayes factor (minimum of the two values computed using AIC and BIC) for model 1 with respect to the other 15 models. All the values are greater than 150, corresponding to strong evidence in favor of model 1 (Raftery, 1995). Estimates were derived for the endogenous and modulatory connection strengths (tables 25.2 and 25.3) of the optimal model and the posterior probabilities that the parameter estimates were greater than zero.

In anatomical terms, effective connectivity could be direct or could occur via a relay, but there must be a structural mediation. Data about anatomical connections between human auditory areas are lacking; there are data showing connections between HG and PT (Tardif & Clarke, 2001), but we are not aware of any data on connections from PT to STS predicted by the model. The basis for the modulatory connection deserves comment. This suggests a change in the connection strength between HG and PT during spectral envelope analysis. This model predicts a selective sensitization of PT to HG afferents that occurs specifically during spectral envelope analysis. In this way, HG acts like a "hidden node" in the system that is not demonstrated in univariate tests for increased activity during spectral envelope abstraction but does have a causal influence on activity in PT during that process.

The model has limitations and should not be regarded as a general synthesis of all aspects of object analysis. Broadly, the existence of a serial model is in accord with the concept of a single pathway for auditory object analysis and supports the concept of PT as a critical "computational hub" (Griffiths & Warren, 2002) at the interface between the abstraction of auditory object properties and further analysis in distinct higher centers for object analysis that also carry out semantic level processing. It should be emphasized that the approach taken here addresses the simplest level of perceptual analysis when the subject is required to attend to the sounds but does not carry out any object-relevant task or semantic level analysis. It will be of considerable interest to examine the effects of task and semantic analysis on connectivity patterns. A number of questions arise, including whether these levels of processing are associated with modulation at later stages of the system (the connection between PT and STS) or additional back connections.

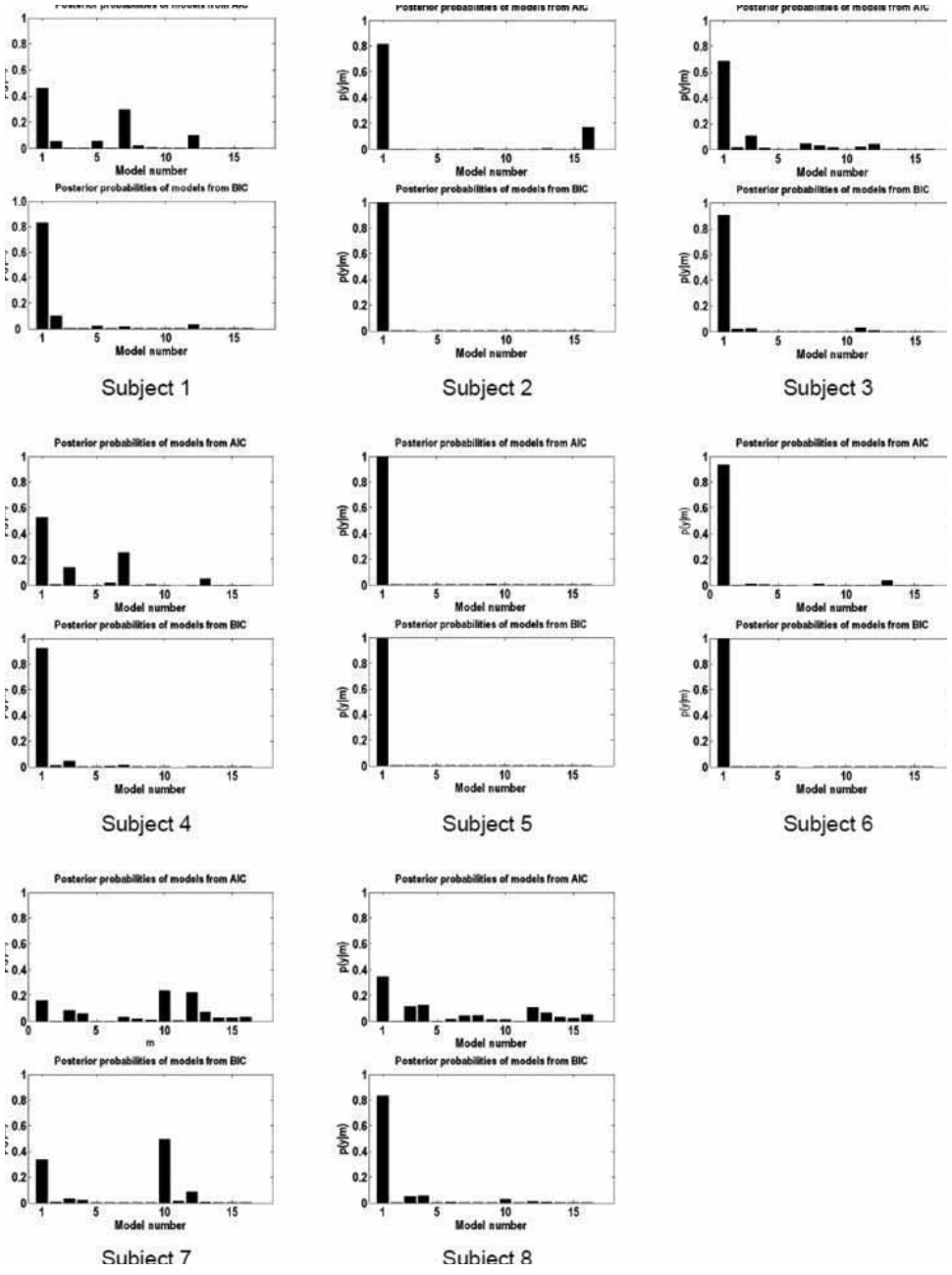


FIGURE 25.11 Evidence for different models. Plots of probabilities $p(y|m)$ for 16 of the models assessed using dynamic causal modeling. The probabilities have been normalized so that they sum to 1. The probabilities represent the probability of the model, given the data, assuming that each model is, a priori, equally likely (Kumar et al., 2007). AIC, Akaike information criterion; BIC, Bayes' information criterion.

Conclusions

The analysis of auditory objects involves a number of different cognitive stages and the term is often used without precision. This synthesis has emphasized functional MRI, but similar approaches can allow inference about stages of object analysis using a number of other techniques, including single-unit recording, EEG, and MEG. The work described here demonstrates that the fundamental level of abstraction of the salient features that define objects, and generalization between them requires specific mechanisms in the superior lateral temporal lobe. Systems identification techniques allow the analysis of specific networks for object perception and have the potential to disambiguate different aspects of auditory object analysis that might use the same computational nodes in different ways.

Table 25.1

Group Bayes' factor for the optimal model (serial model 1)

Model 1 Versus	Group Bayes Factor
2	8.65×10^3
3	4.40×10^2
4	1.06×10^7
5	5.57×10^3
6	7.76×10^4
7	8.03×10^4
8	3.68×10^3
9	4.12×10^2
10	1.69×10^3
11	3.18×10^{30}
12	1.08×10^3
13	2.20×10^2
14	6.61×10^4
15	2.27×10^4
16	9.65×10^3

Table 25.2

Endogenous connection strengths (Hz) for the optimal model (serial model 1)

Subject Number	HG to PT	PT to STS
1	0.54 (0.99)	0.71 (1.0)
2	0.41 (1.0)	0.26 (1.0)
3	0.35 (0.99)	0.41 (1.0)
4	0.49 (1.0)	0.82 (1.0)
5	0.48 (1.0)	0.07 (0.91)
6	0.49 (1.0)	0.52 (1.0)
7	0.10 (1.0)	0.44 (1.0)
8	0.16 (1.0)	0.60 (1.0)

Numbers in parentheses indicate the posterior probability that connection strength is greater than zero.

Table 25.3

Strength of the modulation of the HG → PT connection during spectral envelope analysis

Subject Number	HG to PT	Percent Change
1	0.16 (0.91)	29.4
2	0.54 (1.0)	130
3	0.37 (0.99)	105
4	0.55 (0.99)	113
5	0.26 (1.0)	53.4
6	0.62 (1.0)	125
7	0.13 (0.98)	130
8	0.29 (0.99)	187

Numbers in parentheses indicate the posterior probability that the modulatory strength is greater than zero. The third column is the percent increase in coupling during spectral envelope analysis (the change in the intrinsic connection strength in Table 25.2).

ACKNOWLEDGMENTS T.D.G., S.K., T.O., K.E.S., and K.J.F. are supported by the Wellcome Trust (U.K.) and KvK was supported by the VW Foundation (Germany) and is supported by the Wellcome Trust (U.K.). The ideas developed here were informed by argument with the contributors to a Novartis Foundation discussion meeting on auditory objects in London in 2007. The acosutic size experiments were acquired in collaboration with R. D. Patterson. J. D. Warren (Dementia Research Centre, University College London) acquired the imaging data related to spectral envelope analysis and was involved in the development of a number of the ideas described here.

REFERENCES

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociol. Method. Res.*, *33*, 261–304.

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magn. Reson. Med.*, *39*, 855–864.

Caclin, A., Mcadams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre, space dimensions: A confirmatory study using synthetic tones. *J. Acoust. Soc. Am.*, *118*, 471–482.

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.*, *118*, 887–906.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proc. Biol. Sci.*, *268*, 1669–1675.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302.

Friston, K. J., Mechelli, A., Turner, R., & Price, C. J. (2000). Nonlinear responses in fMRI: The Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage*, *12*, 466–477.

Goebel, R., Roebroeck, A., Kim, D. S., & Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging*, *21*, 1251–1261.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, *61*, 1270–1277.

Griffiths, T. D. (2008). www.staff.ncl.ac.uk/t.d.griffiths/link_teaching_material.htm.

Griffiths, T. D., Bamiou, D.-E., & Warren, J. D. (in press). Disorders of the auditory brain. In A. Rees & A. R. Palmer (Eds.), *The auditory brain*. Oxford, UK: Oxford University Press.

Griffiths, T. D., Buchel, C., Frackowiak, R. S., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nat. Neurosci.*, *1*, 422–427.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci.*, *25*, 348–353.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nat. Rev. Neurosci.*, *5*, 887–892.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cogn. Sci.*, *10*, 14–23.

Harrison, L., Penny, W. D., & Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. *NeuroImage*, *19*, 1477–1491.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.*, *7*, 523–534.

Kawahara, H., & Irino, T. (2004). Underlying prinicples of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In P. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 167–180). Boston: Kluwer Academic.

Kiebel, S. J., David, O., & Friston, K. J. (2006). Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *NeuroImage*, *30*, 1273–1284.

Kikuchi, Y., Horwitz, B., & Mishkin, M. (2007). *Auditory response properties in the rostral and caudal stations of the auditory stimulus processing stream of the macaque superior temporal cortex*. Paper presented at annual meeting of Society for Neuroscience, San Diego.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*, 97–126.

Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., & Griffiths, T. D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Comput. Biol.*, *3*, e100.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150–157.

McIntosh, A., & Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.*, *2*, 2–22.

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., et al. (2002). Neural correlates of timbre change in harmonic sounds. *NeuroImage*, *17*, 1742–1754.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica*, *88*, 320–332.

Moran, R. J., Stephan, K. E., Kiebel, S. J., Rombach, N., O'Connor, W. T., Murphy, K. J., et al. (2008). Bayesian estimation of synaptic physiology from the spectral responses of neural masses. *NeuroImage*, *42*, 272–284.

Patterson, R. D. (2000). Auditory images: How complex sounds are represented in the auditory system. *J. Acoust. Soc. Jpn.*, *21*, 183–190.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*, 767–776.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.*, *24*, 6810–6815.

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004). Comparing dynamic causal models. *NeuroImage*, *22*, 1157–1172.

Raftery, A. (1995). Bayesian model selection in social research. *Sociol. Methodol.*, *25*, 111–196.

Raizada, R. D., & Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, *56*, 726–740.

Russell, B. A. (1945). *History of Western philosophy*. London: Simon & Schuster.

Sawamura, H., Orban, G. A., & Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: A single-cell study of the FMRI adaptation paradigm. *Neuron*, *49*, 307–318.

Schmuckler, M. A., & Gilden, D. L. (1993). Auditory perception of fractal contours. *J. Exp. Psychol.*, *19*, 641–660.

Shamma, S. (2008). On the emergence and awareness of auditory objects. *PLoS Biol.*, *6*, e155.

Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *J. Acoust. Soc. Am.*, *117*, 305–318.

Stephan, K. E. (2004). On the role of general system theory for functional neuroimaging. *J. Anat.*, *205*, 443–470.

Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A., & Friston, K. J. (2007). Comparing hemodynamic models with DCM. *NeuroImage*, *38*, 387–401.

Talairach, P., & Tournoux, J. (1988). *A stereotactic coplanar atlas of the human brain*. Stuttgart, Germany: Thieme.

Tardif, E., & Clarke, S. (2001). Intrinsic connectivity in human auditory areas: Tracing study with DiI. *Eur. J. Neurosci.*, *13*, 1045–1050.

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, *292*, 290–293.

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral thesis. Eindhoven University of Technology, Eindhoven.

Von Kriegstein, K., Smith, D. R., Patterson, R. D., Ives, D. T., & Griffiths, T. D. (2007). Neural representation of auditory size in the human voice and in sounds from other resonant sources. *Curr. Biol.*, *17*, 1123–1128.

Voss, R. F., & Clarke, J. (1975). '1/f noise' in music and speech. *Nature*, *258*, 317–318.

Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *NeuroImage*, *24*, 1052–1057.

Zatorre, R. J., Bouffard, M., & Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *J. Neurosci.*, *24*, 3637–3642.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.*, *14*, 1908–1919.

26 The Cone Photoreceptor Mosaic in Normal and Defective Color Vision

JOSEPH CARROLL, GEUNYOUNG YOON, AND DAVID R. WILLIAMS

ABSTRACT Visual experience is initiated by photons captured in the photoreceptors. The arrangement of these photoreceptors—their topography—limits how we see and sets the stage for how the circuitry in the retina and brain operates. Four different classes of cells are interleaved in the photoreceptor layer of the retina: the rods and three spectral subtypes of cone that form the basis for trichromatic color vision, the long-, middle-, and short-wavelength-sensitive cones (L, M, and S, respectively). A great deal is understood about how the presence of three spectral cone types limits color perception; however, until recently, considerably less had been known about the spatial arrangement of these cone types and how their topography influenced visual experience. Recently, new methods have been developed that enable us to measure the spatial arrangement of cone photoreceptors in the living human eye. These new measurements have produced surprising results that answer some questions about color appearance but raise many others. In this chapter, we review the current understanding of the cone mosaic in normal and defective color vision, emphasizing recent results derived from adaptive optics retinal imaging.

Photoreceptor mosaic in normal color vision

A wealth of histological data are available describing the overall topography of the human photoreceptor mosaic. The most comprehensive data come from Curcio, Sloan, Kalina, and Hendrickson (1990), who showed that while there are gross topographical features of the mosaic that are common across different retinas, there is also considerable variability. For example, as is shown in figure 26.1, the relative rod:cone density varies dramatically across the retina, and this general feature is well preserved in all human retina that have been studied to date. However, Curcio and colleagues (1990) found that the peak foveal cone density varied by at least a factor of 3, though since the data were obtained on postmortem tissue, it was not possible to determine whether such differences had any practical impact on visual function. As we move outward from the center of the fovea (where peak cone density occurs), cone density falls off precipitously until the ora serrata, where there is a significant elevation in cone density (R. W. Williams, 1991).

Using adaptive optics to image the foveal cone mosaic, Putnam and colleagues (2005) observed that the location of peak cone density does not correspond to the preferred retinal locus of fixation; thus there remains ambiguity about the difference between the anatomical fovea and the "functional" fovea. Shown in figure 26.2 are data from three subjects, showing the location of peak cone density with respect to the retinal locus of fixation on individual psychophysical trials. While fixation is in general very accurate, as is shown by the tight cluster of fixation points, there is a systematic deviation from the location of peak cone density. If visual acuity is reciprocally related to cone spacing near the fovea (cf. Green, 1970; Marcos & Navarro, 1997), acuity would have declined by an average of 8% for the subjects in figure 26.2 at the center of fixation compared with the anatomic center of the fovea. This is a relatively small loss in acuity that would be difficult to measure, owing to blurring by the eye's optics, which reduces foveal visual acuity below the cone Nyquist frequency (Marcos & Navarro, 1997). Recent work using adaptive optics to image the cone mosaic of individuals with red-green color vision defects reveals severely disrupted mosaics but normal visual acuity measured with a letter target (Carroll, Neitz, Hofer, Neitz, & Williams, 2004). This further illustrates the insensitivity of standard acuity measures and advocates using interference fringe stimuli that are immune to optical blur to probe the absolute relationship between the cone mosaic and visual acuity.

The human foveal cone mosaic is an efficiently packed mosaic, with the locations of cone centers forming a triangular array. Interleaved within the overall cone mosaic are the three spectral cone submosaics (short-, middle-, and long-wavelength-sensitive; S, M, and L). Since only one spectral type of cone occupies any given location within the cone mosaic, there is an apparent confound as to how the

JOSEPH CARROLL Department of Ophthalmology, Medical College of Wisconsin, Milwaukee, Wisconsin
GEUNYOUNG YOON and DAVID R. WILLIAMS Center for Visual Science, University of Rochester, Rochester, New York

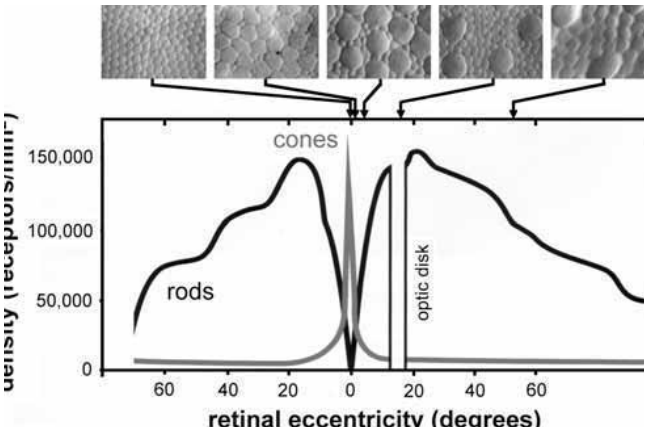


FIGURE 26.1 Nonuniform distribution of rods and cones in the human retina. Plot of photoreceptor density as a function of retinal eccentricity. The top panels show ex vivo images of the photoreceptor mosaic from Curcio, Sloan, Kalina, and Hendrickson (1990). The leftmost image is from the all-cone fovea; the remaining panels contain both rods (smaller cells) and cones (larger cells). While rod density increases dramatically in the peripheral retina, rod diameter remains relatively constant (about 2 μm). Conversely, the cone photoreceptors increase from about 2 μm in diameter at the fovea to about 8 μm at about 10 degrees eccentricity, after which point they remain relatively constant (Samy & Hirsch, 1989). (Modified from Webvision (http://webvision.med.utah.edu), with permission.) (See color plate 33.)

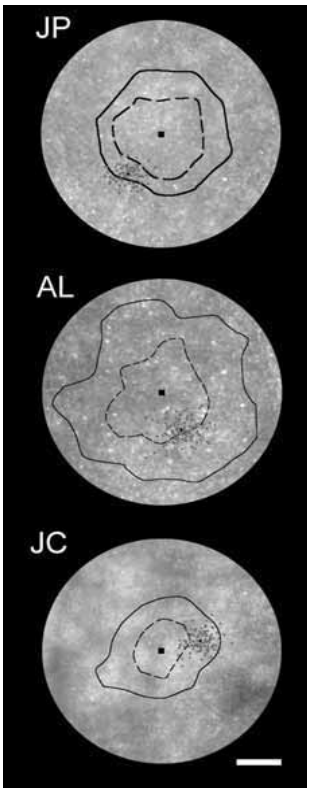


FIGURE 26.2 The area of highest cone density is not always used for fixation. Shown are retinal montages of the foveal cone mosaic for three subjects. The black square represents the foveal center of each subject, as defined by the location of peak cone density. The dashed black line is the isodensity contour line representing a 5% increase in cone spacing, and the solid black line is the isodensity contour line representing a 15% increase in cone spacing. Red dots are individual fixation locations. Scale bar is 50 μm. (Reproduced from Putnam et al., 2005, with permission.) (See color plate 34.)

visual system is able to reliably extract color information at all spatial locations within an image. As such, the precise arrangement of these spectral subtypes has been of great interest, and we discuss the S, M, and L cone mosaics below.

S CONE MOSAIC: STRUCTURAL ORGANIZATION The S cones can easily be distinguished from L and M cones by morphological and histochemical features (Ahnelt & Kolb, 2000; Cornish, Hendrickson, & Provis, 2004; de Monasterio, Schein, & McCrane, 1981; Szel, Diamanstein, & Rohlich, 1988). They are more cylindrical in shape (Curcio et al., 1991), have distinct neural circuitry (Mariani, 1984) and synaptic structure (Ahnelt & Kolb, 2000), and contain a photopigment that is distinct from that found in the L/M cones (Bowmaker & Dartnall, 1980; Nathans, Thomas, & Hogness, 1986). The S cones are relatively sparse throughout the human retina (averaging about 6–8% of the total cone population), with peak density (usually about 10% of the local cone number) occurring near 1-degree eccentricity (Curcio et al., 1991). An interesting feature of the S cone submosaic is that the very central fovea is lacking S cones (König, 1894; Willmer & Wright, 1945; D. R. Williams, MacLeod, & Hayhoe, 1981a, 1981b). The extent of this S cone free zone is about 20 degrees of arc in diameter, though the size and even existence of this area are variable across individuals. In humans, the S cone mosaic is randomly interleaved among the L/M mosaic near the fovea but becomes regularly arranged at more peripheral locations. A number of questions surrounding the S cone mosaic remain, such as why the arrangement of the mosaic is different in nonhuman primates, what the variability is across subjects, and what molecular mechanisms govern the nonuniform placement of S cones within the human retina.

S CONE MOSAIC: FUNCTIONAL CONSEQUENCES It has long been believed that the S cone mosaic is sparser than the other L and M cones because the retinal image quality that is available to the S cones is reduced by the eye's chromatic aberration (Packer & Williams, 2003; D. R. Williams et al., 1981a, 1981b; Yellott, Wandell, & Cornsweet, 1984), although this interpretation is controversial. McLellan, Marcos, Prieto, and Burns (2002) concluded that when monochromatic aberrations are taken into account, chromatic aberration does not, in the average eye, degrade the retinal image quality of the S cones. However, there is theoretical and experimental evidence that the role of

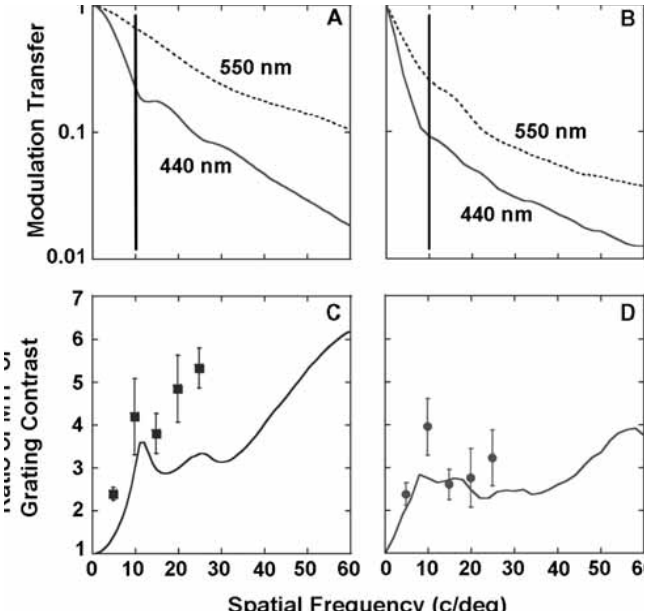


FIGURE 26.3 Chromatic aberration degrades the retinal image quality of the S cone submosaic. (*A*, *B*) Modulation transfer functions at 440 nm (solid line) and 550 nm (dashed line) computed with the measured aberrations for a 3-mm (*A*) and 6-mm (*B*) pupil. Nine wavefront measurements were made on each eye, and the average modulation transfer function (MTF) was calculated. The average root mean square wavefront errors (from second to fifth order without defocus) for the 3- and 6-mm pupil size were 0.2 ± 0.11 μm and 0.82 ± 0.23 μm (mean ± standard deviation), respectively. The amount of defocus was adjusted for each subject to maximize the modulation transfer of a 550-nm grating with a 15 cycle/degree spatial frequency. To compute the MTF of a 440-nm grating, we added 1.0 diopter of defocus induced by longitudinal chromatic aberration when the eye was focused at 550 nm. The Nyquist limit for the S cones is shown as a vertical line at 10 cycles/degree. (*C*, *D*) Solid lines show the ratio of the MTF at 550 nm to that at 440 nm for 3- and 6-mm pupils, respectively. The symbols show matching data from the psychophysical experiment. They represent the ratio of the fixed contrast (50%) of the 440-nm grating to the contrast of the 550-nm grating that matched it in subjective contrast. The error bars in the experimental data represent ±1 standard error.

chromatic aberration does, in fact, substantially degrade the retinal image quality available to S cones.

We calculated modulation transfer functions (MTFs) from the wavefront aberrations of five normal subjects measured with a conventional Shack-Hartmann wavefront sensor. Figures 26.3*A* and 26.3*B* show the average of the horizontal and vertical MTFs. For both pupil sizes, retinal image quality at 550 nm is significantly better than that at 440-nm grating across all spatial frequencies. The difference between the MTFs for 440 nm and 550 nm is smaller for a larger pupil, indicating that the impact of chromatic aberration becomes less dominant when the magnitude of the monochromatic aberrations increases with an increase in pupil size. However, the MTF at 440 nm remains approximately three times on average lower than the 550-nm MTF up to 60 cycles per degree (cpd). Similar conclusions follow from calculations of the overall contrast of the retinal image in white light for each of the three cone classes (Packer & Williams, 2003).

We performed a psychophysical experiment to verify the MTFs. The same subjects matched the perceived contrast of two gratings of equal luminance. In each half of a 1.5-degree visual field, a sinusoidal grating with the same spatial frequency was simultaneously viewed in a split field with 440-nm and 550-nm light with paralyzed accommodation. The subject maximized contrast of a 15 cpd sinusoidal grating in the 550-nm half of the field. Contrast matching of the 550-nm grating to the 440-nm grating with 50% contrast was subjectively performed at each of five different spatial frequencies presented in random order. At each experimental condition, four trials were conducted for both horizontal and vertical gratings and were averaged.

Figures 26.3*C* and 26.3*D* show the ratios of the MTF at 550 nm to those at 440 nm and the contrast of the 440-nm grating to the matched contrast of the 550-nm grating. If the contrast of the 440-nm grating were decreased by chromatic aberration, this ratio would be larger than 1 because subjects would need to reduce the contrast of the 550-nm grating to match that of the 440-nm grating blurred by chromatic aberration. In agreement with the ratios computed from the MTFs, the subjective contrast of the 440-nm grating was decreased by a factor of approximately 3 (6-mm pupil) and 4 (3-mm pupil) on average compared to the 550-nm grating. All five subjects reported a reduction in subjective contrast of the 440-nm grating at all spatial frequencies. Because 440-nm and 550-nm gratings differ in color, it was conceivable that neural factors, rather than chromatic aberration, were responsible for the selective subjective blur of short-wavelength gratings. Therefore the experiment was repeated in two subjects with 540-nm and 420-nm gratings chosen to minimize differential activation of L and M cones. Moreover, differential S cone contribution was greatly reduced by superimposing the gratings on a steady violet background. With this method, both sides of the field appeared blue, with almost no discernable color difference. Nonetheless, a large reduction in contrast of the 420-nm grating relative to the 540-nm grating persisted, consistent with a largely optical rather than a neural effect. Our experimental and theoretical results both confirm that the eye's chromatic aberration causes a substantial decrease in retinal image quality even when the eye's wavefront aberration contains large amounts of higher-order aberrations. Contrary to the conclusions of McLellan and colleagues (2002), the blur produced by chromatic aberration is substantial despite the diluting effect of the monochromatic aberrations, consistent with the widely accepted explanation for the sparseness of the S cone submosaic. At 1-degree eccentricity, the S cone mosaic has a Nyquist limit that is approximately 4.1 times less than that

of the M/L cones (Curcio et al., 1991; Hofer, Carroll, Neitz, Neitz, & Williams, 2005; Roorda & Williams, 1999). For a 3-mm pupil, the spatial bandwidth of the 550-nm MTF is about 3.2 times greater than that for the 440-nm MTF, showing that the relationship between optical quality and sampling is not very different for the S cones and the M and L cones.

L and M Cone Mosaic: Structural Organization There is no antibody that can distinguish L cones from M cones; this is because the photopigments they contain are 96% identical (Nathans, Thomas, & Hogness, 1986). As such, much of what we know about the L/M submosaic has come from indirect measurements of the retina. Recent direct work using adaptive optics and molecular analyses has for the most part confirmed previous results, though it has uncovered surprising levels of variation within this mosaic.

Given that they make up about 95% of the total cone population and thus drive the majority of our visual activity, there has been considerable interest in the L and M cones, both in their relative numbers (Carroll, Neitz, & Neitz, 2002; DeVries, 1946; Dobkins, Thiele, & Albright, 2000; Jacobs & Neitz, 1993; Kremers et al., 2000; Pokorny, Smith, & Wesner, 1991; Roorda & Williams, 1999; Rushton & Baker, 1964) and in their topographical arrangement throughout the mosaic (Balding, Sjoberg, Neitz, & Neitz, 1998; Bowmaker, Parry, & Mollon, 2003; Deeb, Diller, Williams, & Dacey, 2000; Hagstrom, Neitz, & Neitz, 1998; Knau, Jägle, & Sharpe, 2001; Mollon & Bowmaker, 1992; Packer, Williams, & Bensinger, 1996; Roorda, Metha, Lennie, & Williams, 2001). For many years, scientists used indirect measures to assess the homogeneity of these ratios across the retina and across observers. With the advent of adaptive optics, it has become possible to examine the L/M mosaic noninvasively and directly. What is now clear is that in humans with normal color vision, there are on average two L cones for every M cone, there is variability in the relative numbers of L and M cones between people, and the ratio of L to M cones is not completely uniform across an individual retina.

While most studies suggest that the ratio of L to M cones is probably constant across the central retina, the evidence for this is somewhat inferential. In fact, direct evidence from adaptive optics and retinal densitometry has shown that in at least one individual, the relative numerosity of L and M cones is not homogenous across the central retina (Hofer et al., 2005). However, there are other, larger-scale inhomogeneities in the primate L/M mosaic. There is a nasal-temporal asymmetry in the local L-to-M cone ratio in macaque retina, though whether this asymmetry is a prominent feature of human retinae is not clear. Data from mRNA analysis and cone isolating mfERG of human retina reveal that the relative L-to-M cone ratio increases significantly in the peripheral retina, reaching nearly an all L cone mosaic at the edge of the retina (M. Neitz, Balding, McMahon, Sjoberg, & Neitz, 2006). Shown in figure 26.4 is a topographical map of L-to-M mRNA, revealing the dramatic and systematic variation across the retina.

As was mentioned above, numerous indirect studies have suggested that there are on average about twice as many L cones as M in the human retina, with large intersubject variability. Direct information on the numbers and locations of L, M, and S cones obtained with spatially localized retinal densitometry in 10 living human subjects reveals the extent of this variability in L-to-M cone ratio across individuals with normal color vision to be over a 40-fold range (Hofer et al., 2005). Shown in figure 26.5 are pseudo-colored images of the human cone mosaic, showing the remarkable variation in L-to-M cone ratio. Also evident from these images is the fact that the L and M cone submosaics are randomly interleaved.

L and M Cone Mosaic: Functional Consequences In the face of the dramatic intersubject and intrasubject variation in the L-to-M cone mosaic, questions arise regarding the behavioral consequence of such variability. The fundamental experiment in color vision, color matching with spatially uniform fields, is very sensitive to the spectral absorptance of the cone photopigments but invariant with respect to the local ratio of cone types. Thus both across observers and across visual field position for a single observer, color matches will be preserved despite variations in the ratios of cone types. There are pronounced deficits in L/M color vision in the periphery compared to the central retina (Gordon & Abramov, 1977; Mullen, 1991); however, this has more to do with postreceptoral sampling than with L/M numerosity. Even the variation in L-to-M ratio between subjects has been shown to have little consequence for color vision, despite previous hypotheses (cf. Cicerone, 1987). For example, Miyahara, Pokorny, Smith, Baron, and Baron (1998) showed that in two female carriers of a red/green color vision defect, despite a dramatic skew in their L-to-M cone ratio, their red/green color vision was completely normal. J. Neitz, Carroll, Yamauchi, Neitz, and Williams (2002) used the flicker photometric ERG to probe L-to-M ratio in over 60 individuals and showed that that wavelength of unique yellow (the presumed null point of the red-green system) did not change across subjects, suggesting a postreceptoral normalization mechanism that compensates for any biases in L-to-M cone ratio. The reality is that the human visual system is quite resilient to variation at the retinal level, though this is obviously dependent on the sensitivity of the test used to probe color vision.

The fact that the three cone submosiacs each sample the retinal image at a lower rate than the overall mosaic makes the retina susceptible to aliasing at lower spatial frequencies

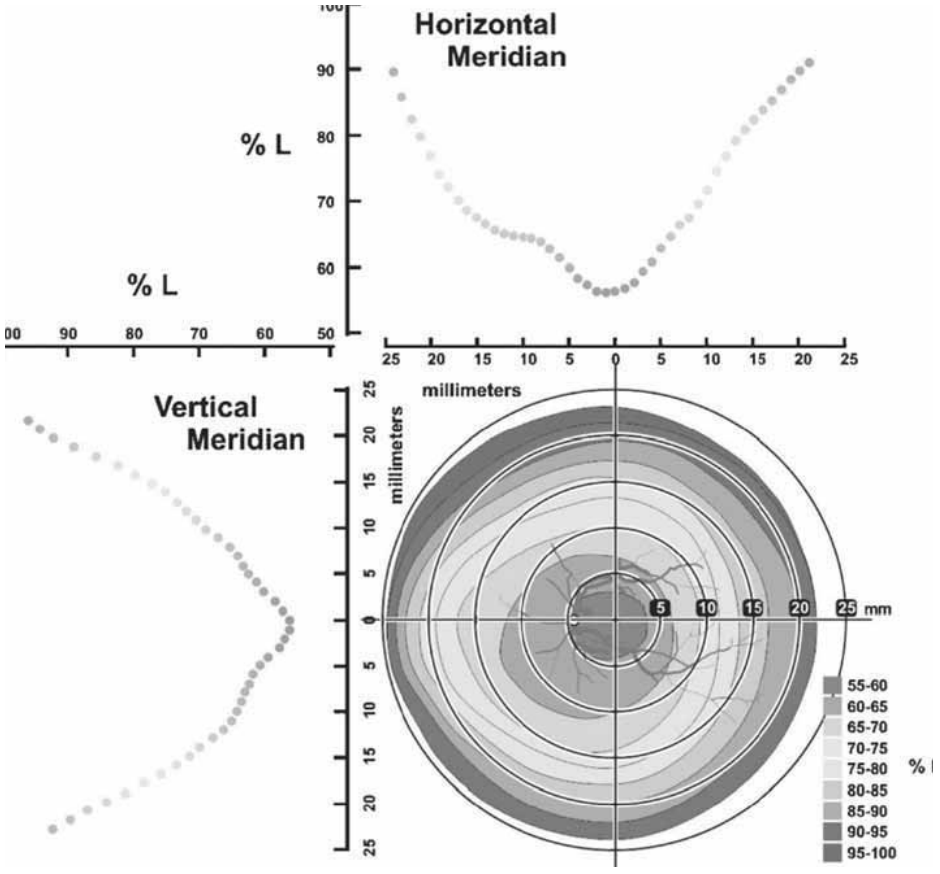


FIGURE 26.4 The L-to-M cone ratio is not constant across the retina. The lower right panel shows a topographical map of the percent of L-opsin mRNA in a human donor retina. The proportion of L-opsin to M-opsin mRNA is directly related to the relative numbers of L and M cones at a locus, assuming that L and M cones produce the same absolute amounts of mRNA. Horizontal and vertical meridian slices show the dramatic increase in percent of L as a function of eccentricity. (Reproduced from Neitz et al., 2006, with permission.) (See color plate 35.)

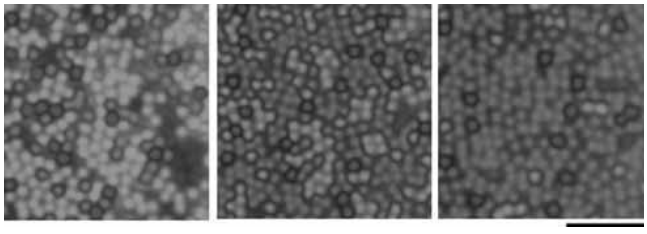


FIGURE 26.5 Intersubject variation in L-to-M cone ratio. False color images showing the arrangement of L (red), M (green), and S (blue) cones in the retinas for three human subjects. The identity of each cone as L, M, or S was inferred from retinal densitometry measurements obtained with an adaptive optics fundus camera. The proportion of S cones does not vary significantly between subjects; however, the L-to-M cone ratio can vary by a factor of 40 across individuals with normal color vision. Scale bar is 5 arcmin. (See color plate 36.)

than would be expected from a monochrome mosaic with the same total number of cones. The aliases that are produced by the cone submosaics exposed to luminance patterns would be expected to be strongly chromatic. It is possible to observe submosaic aliasing with high-frequency achromatic gratings seen through the S cones (D. R. Williams & Collier, 1983) and the L and M cones (D. R. Williams, Sekiguchi, Haake, Brainard, & Packer, 1991). However, submosaic aliasing is fleeting and subtle in visual experience, contrary to the predictions of simple sampling and interpolation models. This suggests that the visual system must have a more sophisticated mechanism to expunge submosaic aliasing from visual experience. Brainard, Williams, and Hofer (2008) have suggested that well-known features of human color and spatial vision may play a role in reducing submosaic aliasing. Specifically, the spatial bandwidth of postreceptoral chromatic mechanisms is lower than that of the achromatic mechanism (Mullen, 1985; Poirson & Wandell, 1993), and this difference persists even when optical blurring and chromatic aberration are eliminated with laser interferometry (Sekiguchi, Williams, & Brainard, 1993a, 1993b). This difference reduces submosaic aliasing by suppressing chromatic interpretations of high spatial frequencies in the retinal image. There is additional evidence for spatial averaging of color information with suprathreshold stimuli (Cao & Shevell, 2005), collectively referred to as

assimilation, that is qualitatively similar to the reduced spatial bandwidth for chromatic mechanisms assessed at detection threshold. It is possible that assimilation effects occur because of the need to protect against submosaic aliasing, though a quantitative model of the benefits assimilation could provide has yet to be made.

Cone topography in inherited color vision deficiencies

There are a number of instances in which normal color discrimination is impaired, and there is now a detailed understanding of the molecular mechanisms underlying these defects. Nearly all inherited color vision defects have their origin in a disruption of normal cone photopigment expression; either a cone pigment is absent or a mutant pigment is expressed. Until recently, little attention had been given to the residual photoreceptor mosaic of these individuals. Results using adaptive optics retinal imaging have stimulated a reevaluation of the ideas about what the appearance of these mosaics might be. Here, we review the four major types of inherited color deficiencies and discuss what has been revealed about the accompanying cone photoreceptor mosaic.

Tritanopia Tritan color vision deficiency is an inherited autosomal dominant abnormality of S cone function (Wright, 1952). The disorder is reported to exhibit incomplete penetrance, meaning that individuals with the same underlying mutation manifest different degrees of color vision impairment (Cole, Henry, & Nathan, 1966; Kalmus, 1955; Miyake, Yagasaki, & Ichikawa, 1985; Pokorny, Smith, & Went, 1981; Went & Pronk, 1985). Four different amino acid substitutions in the S cone photopigment have been associated with tritanopia, which is only slightly more rare than autosomal dominant retinitis pigmentosa (adRP) (Gunther, Neitz, & Neitz, 2006; Weitz et al., 1992; Weitz, Went, & Nathans, 1992). Each substitution occurs at an amino acid position that lies in one of the transmembrane alpha helices of the protein and is therefore expected to interfere with folding, processing, or stability of the encoded opsin.

In adRP, rod photopigment mutations are associated with degeneration of the associated photoreceptors. This is due in part to the fact that the photopigment plays such an important structural role in maintaining the integrity of the outer segment, comprising nearly 90% of the protein in the outer segment. To investigate whether there is S cone degeneration in autosomal dominant tritan defects, Baraas and colleagues (2007) examined two related tritan subjects (a 57-year-old male and his 34-year-old daughter) both heterozygous for a novel mutation in their S-opsin gene. The mutation resulted in a substitution of glutamine for arginine at position 283 (R283Q). Given the sensitive microenvironment of the photopigment, the substitution of a polar, neutral amino acid for a positively charged one would be expected to compromise the function of the photopigment. The father manifests as tritanopic on all color vision tests, whereas the daughter made mild tritan errors on only a small subset of color vision tests. Interestingly, the father reports that it has only been in recent years that he has noticed difficulties with discriminating between some colors such as orange-yellow and pink, while the daughter reports never having any color discrimination problems. We used adaptive optics ophthalmoscopy to obtain high-resolution images of the cone mosaic of both individuals combined with retinal densitometry to identify S cones in the mosaic. Surprisingly, while normal S cone density was reported for the daughter (4.9%, or 2224 cones/mm^2), no evidence for S cones was observed in the father, though the overall cone density was within normal limits for both individuals. Since S cones normally occupy a small minority of the total cone population and since cone density is so variable across normal individuals (Curcio et al., 1990; Gao & Hollyfield, 1992), it is not surprising that the cone density of the father appeared normal despite the apparent absence of S cones.

One feature of the cone mosaic that *can* be exploited to study subtle disruptions in the packing geometry of the cone mosaic is the spatial regularity. The spatial regularity of the mosaic can be assessed by using a number of metrics (Cook, 1996; Rodieck, 1991); one of the more intuitive ones is the Voronoi analysis (Curcio & Sloan, 1992; Pum, Ahnelt, & Grasl, 1990). With this analysis, individual cones are represented as points in a two-dimensional plane. For each cell, a Voronoi domain is constructed by defining points in the plane that are closer to that cell than any other cell in the mosaic. The number of sides of the resultant polygon reflects the packing geometry of the local mosaic. In a perfectly regular mosaic with triangular packing, each cell would have a hexagonal Voronoi domain. Shown in figure 26.6*A* is the Voronoi analysis of a normal human cone mosaic and that of the tritan father mosaic. The polygons are color coded according to the number of sides they have, with green indicating six sides. While the majorities are green (indicating a largely triangular mosaic), there are many fractures in the regularity of the mosaic. These disruptions have been hypothesized to correlate with the location of S cones in the mosaic (Pum et al., 1990); however, compelling evidence for this is lacking. Nevertheless, the father's mosaic was significantly more irregular than normal (figure 26.6*C*), while the daughter's mosaic was indistinguishable from normal (figure 26.6*B*). The disparate S cone mosaics in these two subjects are consistent with their distinct behavioral phenotypes, the increased irregularity in the father's mosaic likely being a remnant of the degeneration of the S cones in his mosaic.

The work reported by Baraas and colleagues (2007) provides the first anatomical evidence that tritan phenotypes

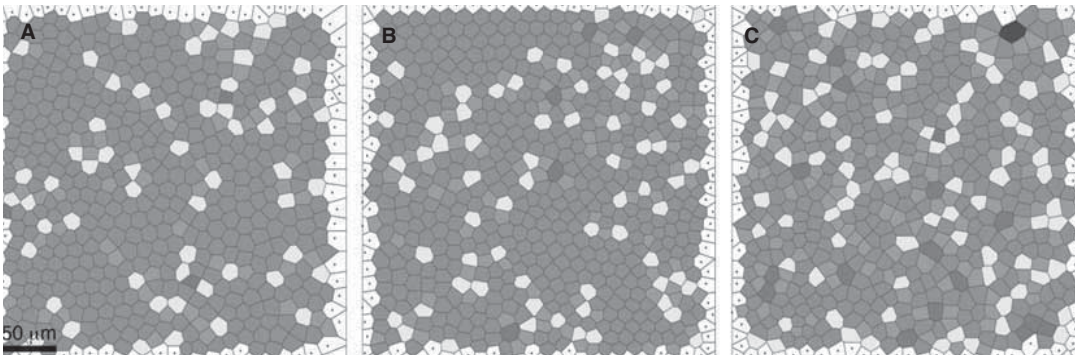


FIGURE 26.6 Regularity of the human cone mosaic. Voronoi domain associated with each cone photoreceptor in a patch of retina from (*A*) a normal trichromat, (*B*) a 34-year-old female with a mild tritan defect, and (*C*) a 57-year-old male with a severe tritan defect. The color code indicates the number of sides on each Voronoi polygon (magenta = 4, cyan = 5, green = 6, yellow = 7, red = 8, purple = 9). Large regions of six-sided polygons indicate a regular triangular lattice, whereas other colors mark points of disruptions in the hexagonal packing of the foveal mosaic. Despite the fact that the father and the daughter carried the same heterozygous mutation in their S-opsin genes (predicting a tritan phenotype), the regularity of the father's mosaic was significantly disrupted, while the daughter's was indistinguishable from normal. Scale bar is 50 μm. (Reproduced from Baraas et al., 2007, with permission.) (See color plate 37.)

associated with S-opsin mutations can be associated with the loss of S cones. This suggests a mechanism in which the mutations produce their effects by reducing the viability of the S cones (similar to the mechanism in adRP), and heterozygotes that express both the normal and mutant S opsins will exhibit trichromatic color vision that can be indistinguishable from normal until the S cones succumb to the toxicity of the mutant opsin.

RED/GREEN COLOR VISION DEFICIENCY The most common form of inherited color vision deficiency is one that affects the red-green (L-M cone) system. Among individuals of Western European ancestry, about 7–10% of males have a red-green color vision defect. The incidence in females is much lower (approximately 0.4%) because the defects are inherited as X-linked recessive traits, though approximately 15% of females are carriers of a red-green defect. The general genetic causes of red-green color vision deficiency involve a disruption of the L/M gene array on the X-chromosome. The most common cause is rearrangement of the L/M genes resulting either in the deletion of all but one visual pigment gene or in the production of a gene array in which the first two genes both encode a pigment of the same spectral class (Deeb et al., 1992; Jagla, Jägle, Hayashi, Sharpe, & Deeb, 2002; Nathans, Piantanida, Eddy, Shows, & Hogness, 1986; M. Neitz et al., 2004; Ueyama et al., 2003). The second general cause is the introduction of an inactivating mutation in either the first or second gene in the array. The most prevalent inactivating mutation results in the substitution of arginine for cysteine at position 203 (C203R) in the L/M pigment (Bollinger, Bialozynski, Neitz, & Neitz, 2001; M. Neitz et al., 2004; Winderickx et al., 1992). Cysteine 203 forms an essential disulfide bond being highly conserved among G-protein-coupled receptors (Sakmar, 2002). This mutation was first observed in blue cone monochromacy (Nathans et al., 1989) where it was shown to directly disrupt photopigment function (Kazmi, Sakmar, & Ostrer, 1997). Mutating the corresponding cysteine residue in human rhodopsin (position 187) causes autosomal dominant retinitis pigmentosa (Richards, Scott, & Sieving, 1995).

The two different causes of red/green color vision defects might be expected to have different retinal phenotypes. It is thought that all photoreceptors that are destined to become L or M cones will express either the first or second gene in the X-chromosome array (Hayashi, Motulsky, & Deeb, 1999). In the case of gene rearrangements, all photoreceptors are expected to express a gene that encodes a functional pigment, though these would all be of the same spectral type. However, in the case of inactivating mutations, a fraction of the photoreceptors will express a pigment that is not functional and, in fact, may be deleterious to the viability of the cell. Recently, it was discovered that there are different retinal phenotypes among red-green color-blind individuals. Carroll, Porter, Neitz, Williams, and Neitz (2005) found that in individuals having either a single-gene array or an array in which the first two genes both encode a pigment of the same spectral class, the cone mosaic is normal in appearance. In contrast, in individuals in whom one of the genes in the array encodes a pigment with an inactivating mutation, dramatic loss of healthy cones is observed, consistent with the hypothesis that cells expressing the mutant pigment degenerated (Carroll et al., 2004). Shown in figure 26.7 are adaptive optics images from individuals with color vision defects caused by photopigment mutations.

Besides a reduction in color discrimination, the disrupted mosaics (sometimes having 60% fewer cones than normal) might also be expected to confer a reduction in spatial vision.

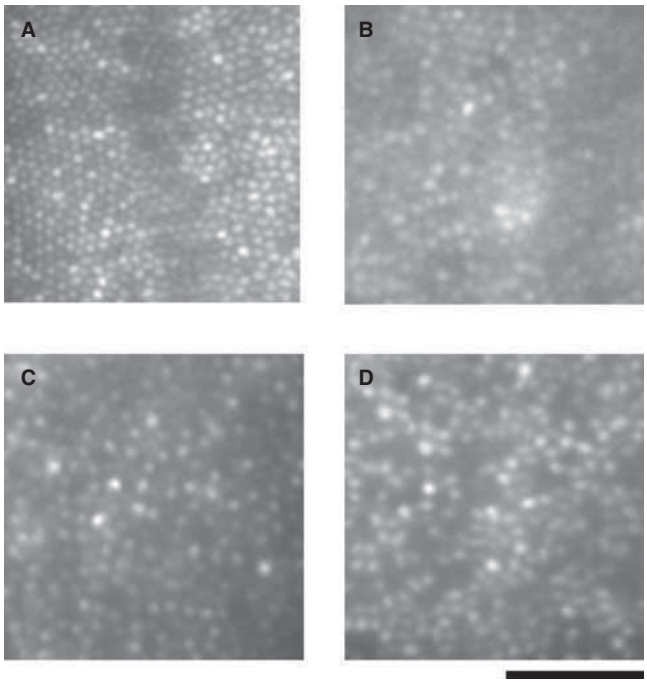


FIGURE 26.7 Images of the cone mosaic from individuals with different cone opsin mutations. Images are from 1-degree temporal retina from (*A*) a normal trichromat, (*B*) a blue-cone monochromat carrier (Carroll, Porter, Neitz, Williams, & Neitz, 2005), (*C*) a protanope male (Carroll et al., 2007), and (*D*) a deuteranope male (Carroll, Neitz, Hofer, Neitz, & Williams, 2004). The density of cones in the BCM carrier (*B*) is reduced significantly (21,684 cones/mm^2 for the BCM carrier retina, compared with an average normal density at this retinal eccentricity of 55,184 cones/mm^2), though the mosaic appears continuously packed (the cells are slightly larger in diameter). The density of the cones in the males with red/green defects is also reduced, though the appearance of the mosaic is quite different, with numerous punctate gaps in the cone mosaic. Sale bar is 50 μm.

However, clinical tests of visual acuity and visual field sensitivity are completely normal in these individuals. To demonstrate that the dark areas in the retinal images indeed correspond to nonfunctional cones, Makous and colleagues (2006) used adaptive optics to deliver small stimuli to the retina. By using adaptive optics to correct for the eye's aberrations, it is possible to confine a majority of the energy of the spot to an area of retina about the size of a single cone aperture. They found a marked reduction in sensitivity that was almost exactly proportional to the amount of retinal mosaic occupied by dark spaces. Thus the absence of a clinical deficit likely reflects redundancy in the stimuli used in the diagnostic tests, which causes correlated responses even in neurons whose sensitivities do not overlap. For example, the smallest spots used in clinical perimetry subtend 0.11 degrees. Even without optical spread, the image of such a spot covers some 150 cones at the center of the fovea of an average retina. Nevertheless, the fact that a visual deficit is apparent only when bypassing the eye's optical aberrations further highlights the limiting effect these aberrations can have on our normal visual activity.

BLUE CONE MONOCHROMACY Blue cone monochromacy (BCM) is a condition in which L and M cone function is absent (Pokorny, Smith, & Verriest, 1979). This is caused by either a deletion of essential DNA elements needed for normal transcription of the pigment genes or a deletion of all but one of the X-chromosome visual pigment genes with the one remaining gene containing a missense mutation (Ayyagari et al., 2000; Nathans et al., 1989, 1993). Affected individuals have very poor acuity, myopia, nystagmus, poor color discrimination, and minimally detectable ERG responses.

Owing to the X-linked nature of the condition, female carriers are spared from a full manifestation of the associated defects but can show abnormal cone ERG amplitudes (Berson, Sandberg, Maguire, Bromley, & Roderick, 1986). Using adaptive optics, Carroll and colleagues (2005) imaged the photoreceptor mosaic of a BCM carrier and found a mosaic that had 60% fewer cones than normal, with foveal cones that were larger in diameter than normal (see figure 26.7). As a result of X-inactivation, on average half of the L/M cones will express a gene from a normal L/M array, while the other half will attempt to express a gene from the array harboring the BCM mutation. The loss of cones is believed to reflect the fact that those cones that were unable to express a photopigment gene degenerated early in development. The preliminary observations in the BCM carrier predict that an affected male would have a completely disrupted cone mosaic, with only about 5–10% of the normal cone complement. No adaptive optics data on BCM males are available; however, indirect data from optical coherence tomography suggests that there is an alteration of the outer segment of the photoreceptors and an alteration in the area of the somata of the photoreceptors in males with BCM (Barthelmes et al., 2006). Future work to compare these complementary imaging modalities should yield valuable information about the origin of the disrupted OCT signal in these BCM males.

ROD MONOCHROMACY Rod monochromacy is a rare disorder (approximately 1 in 30,000) typically characterized by a lack of color discrimination, photophobia, reduced acuity, visual nystagmus, and nondetectable cone electroretinograms (see Hess, Sharpe, & Nordby, 1990, for thorough reviews). Complete achromatopsia has been linked to numerous mutations in CNGA3 and CNGB3 (which encode the α- and β-subunits, respectively of the cone cyclic-nucleotide gated channel), as well as GNAT2, which encodes the α-subunit of the cone G-protein transducin. To account for the apparent absence of cone function, Galezowski (1868) first proposed a rod-only theory in which the cones of the

retina are malformed or completely absent and visual function takes place entirely in the rod photoreceptors. Subsequent histological studies partially discounted this theory, but the result is an inconsistent picture of the photoreceptor mosaic of the achromat. Larsen (1921) reported scarce, malformed retinal cones in the fovea and normal peripheral cones in the retina of a 29-year-old female. Harrison, Hoefnagel, and Hayward (1960) found imperfectly shaped and reduced numbers of cones throughout the retina of a 19-year-old male. In the retina of a 69-year-old female, Falls, Wolter, and Alpern (1965) observed normal numbers of odd-shaped foveal cones and scarce numbers in the periphery. Most recently, Glickstein and Heath (1975) found no evidence of cones in the fovea and reduced numbers of cones elsewhere in the retina of an 85-year-old male.

Nishiguchi, Sandberg, Gorji, Berson, and Dryja (2005) found significant phenotypic variation among rod monochromats—specifically that some had small amounts of remaining cone function. It seems likely that this would be at least partly driven by the underlying genetic variation. As a first step toward developing a genotype-phenotype map in rod monochromacy, Carroll, Choi, and Williams (2008) obtained the first images of the photoreceptor mosaic from a living rod monochromat for whom the genetic cause of the disease was known. As is shown in figure 26.8, clear images of the photoreceptor mosaic were obtained; however, they were not normal in appearance. The rod monochromat had a disrupted photoreceptor mosaic, and the cells in the mosaic were comparable to rod diameters, not cone diameters, consistent with the hypothesis that this retina is largely (if not completely) devoid of healthy cone photoreceptors. While the appearance of the photoreceptor mosaic in rod monochromacy might seem trivial, given that new gene therapies are on the horizon that promise restoration of cone function in rod monochromats (Alexander et al., 2007; Komaromy et al., 2007), it seems especially relevant to consider the photoreceptor substrate with which any such therapy would have to work. Even beyond the photoreceptors, it has been shown that the retinal circuitry remodels in response to photoreceptor degeneration (see Jones & Marc, 2005, for a

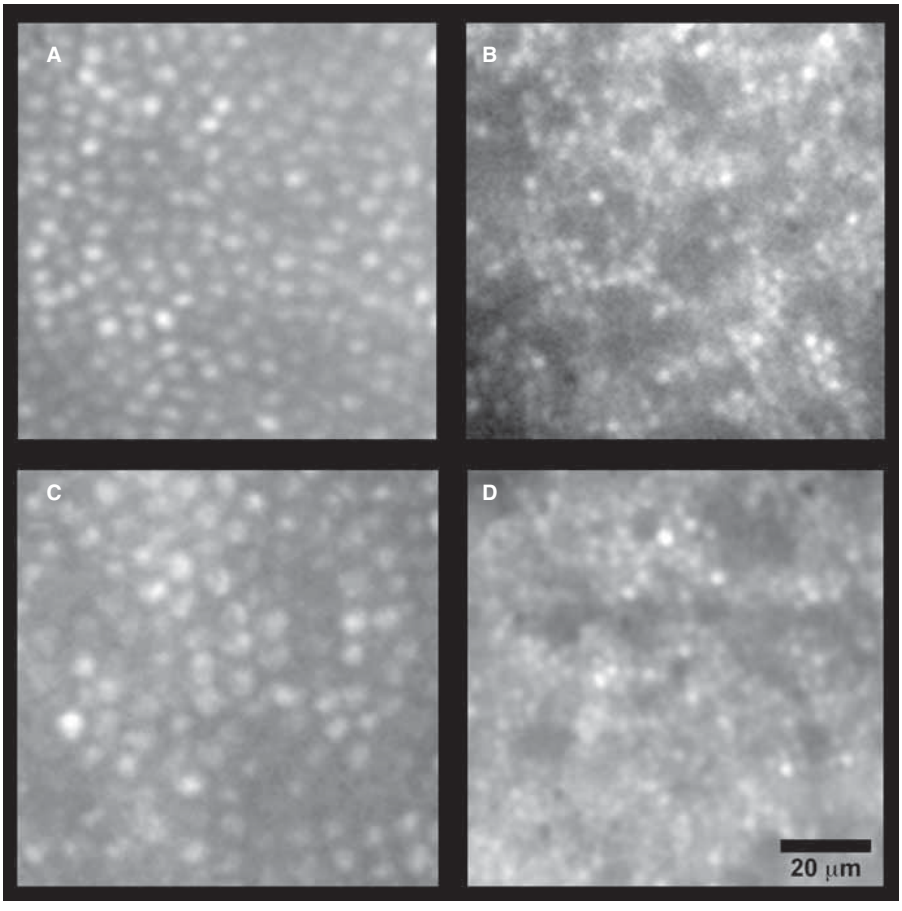


FIGURE 26.8 The retina of the rod monochromat is highly unusual. (*A*, *C*) Images from a 28-year-old normal male, who had been imaged as part of a number of unrelated studies over the course of 3 months. (*B*, *D*) Images from a 28-year-old rod monochromat. Images are from 2.5 degrees (*A*, *B*) and 4 degrees (*C*, *D*) temporal retina. The size and density of the visible cells were typical for rod, not cone, photoreceptors. Scale bar is 20 μm. (Reproduced from Carroll, Choi, & Williams 2008, with permission.)

review) and that cortical signals in rod monochromats are abnormal (Baseler et al., 2002). Thus gene therapies for rod monochromacy would also need to account for the developmental reorganization in the visual cortex, as well as any potential remodeling of the retinal circuitry.

REFERENCES

Ahnelt, P. K., & Kolb, H. (2000). The mammalian photoreceptor mosaic-adaptive design. *Prog. Retinal Eye Res.*, *19*(6), 711–777.

Alexander, J. J., Umino, Y., Everhart, D., Chang, B., Min, S. H., Li, Q., et al. (2007). Restoration of cone vision in a mouse model of achromatopsia. *Nat. Med.*, *13*(6), 685–687.

Ayyagari, R., Kakuk, L. E., Bingham, E. L., Szczesny, J. J., Kemp, J. A., Toda, Y., et al. (2000). Spectrum of color gene deletions and phenotype in patients with blue cone monochromacy. *Hum. Genet.*, *107*, 75–82.

Balding, S. D., Sjoberg, S. A., Neitz, J., & Neitz, M. (1998). Pigment gene expression in protan color vision defects. *Vis. Res.*, *38*(21), 3359–3364.

Baraas, R. C., Carroll, J., Gunther, K. L., Chung, M., Williams, D. R., Foster, D. H., et al. (2007). Adaptive optics retinal imaging reveals S-cone dystrophy in tritan color-vision deficiency. *J. Opt. Soc. Am. [A]*, *24*(5), 1438–1446.

Barthelmes, C., Sutter, F. K., Kurz-Levin, M. M., Bosch, M. M., Helbig, H., Niemeyer, G., et al. (2006). Qualitative analysis of OCT characteristics in patients with achromatopsia and blue-cone monochromatism. *Invest. Ophthalmol. Vis. Sci.*, *47*(3), 1161–1166.

Baseler, H. A., Brewer, A. A., Sharpe, L. T., Morland, A. B., Jägle, H., & Wandell, B. A. (2002). Reorganization of human cortical maps caused by inherited photoreceptor anomalies. *Nat. Neurosci.*, *5*, 364–370.

Berson, E. L., Sandberg, M. A., Maguire, A., Bromley, W. C., & Roderick, T. H. (1986). Electroretinograms in carriers of blue cone monochromatism. *Am. J. Ophthalmol.*, *102*(2), 254–261.

Bollinger, K., Bialozynski, C., Neitz, J., & Neitz, M. (2001). The importance of deleterious mutations of M pigment genes as a cause of color vision defects. *Color Res. Appl.*, *26*, S100–S105.

Bowmaker, J. K., & Dartnall, H. J. A. (1980). Visual pigments of rods and cones in a human retina. *J. Physiol. Lond.*, *298*, 501–511.

Bowmaker, J. K., Parry, J. W. L., & Mollon, J. D. (2003). The arrangement of L and M cones in human and a primate retina. In J. D. Mollon, J. Pokorny & K. Knoblauch (Eds.), *Normal and defective colour vision* (pp. 39–50). New York: Oxford University Press.

Brainard, D. H., Williams, D. R., & Hofer, H. (2008). Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color appearance of small spots. *J. Vis.*, *8*(5), 1–23.

Cao, D., & Shevell, S. K. (2005). Chromatic assimilation: Spread light or neural mechanism? *Vis. Res.*, *45*(8), 1031–1045.

Carroll, J., Choi, S. S., & Williams, D. R. (2008). In vivo imaging of the photoreceptor mosaic of a rod monochromat. *Vis. Res.*, *48*(26), 2564–2568.

Carroll, J., Neitz, M., Hofer, H., Neitz, J., & Williams, D. R. (2004). Functional photoreceptor loss revealed with adaptive optics: An alternate cause for color blindness. *Proc. Natl. Acad. Sci. USA*, *101*(22), 8461–8466.

Carroll, J., Neitz, M., & Neitz, J. (2002). Estimates of L:M cone ratio from ERG flicker photometry and genetics. *J. Vis.*, *2*(8), 531–542.

Carroll, J., Porter, J., Neitz, J., Williams, D. R., & Neitz, M. (2005). Adaptive optics imaging reveals effects of human cone opsin gene disruption. *Invest. Ophthalmol. Vis. Sci.*, *46*, ARVO E-Abstract 4564.

Cicerone, C. M. (1987). Constraints placed on color vision models by the relative numbers of different cone classes in human fovea centralis. *Die Farbe*, *34*, 59–66.

Cole, B. L., Henry, G. H., & Nathan, J. (1966). Phenotypical variations of tritanopia. *Vis. Res.*, *6*, 303–313.

Cook, J. E. (1996). Spatial properties of retinal mosaics: An empirical evaluation of some existing measures. *Vis. Neurosci.*, *13*(1), 15–30.

Cornish, E. E., Hendrickson, A. E., & Provis, J. M. (2004). Distribution of short-wavelength-sensitive cones in human fetal and postnatal retina: Early development of spatial order and density profiles. *Vis. Res.*, *44*, 2019–2026.

Curcio, C. A., Allen, K. A., Sloan, K. R., Lerea, C. L., Hurley, J. B., Klock, I. B., et al. (1991). Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *J. Comp. Neurol.*, *312*, 610–624.

Curcio, C. A., & Sloan, K. R. (1992). Packing geometry of human cone photoreceptors: Variation with eccentricity and evidence for local anisotropy. *Vis. Neurosci.*, *9*, 169–180.

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *J. Comp. Neurol.*, *292*, 497–523.

de Monasterio, F. M., Schein, S. J., & McCrane, E. P. (1981). Staining of blue-sensitive cones of the macaque retina by a fluorescent dye. *Science*, *213*, 1278–1281.

Deeb, S. S., Diller, L. C., Williams, D. R., & Dacey, D. M. (2000). Interindividual and topographical variation of L:M cone ratios in monkey retinas. *J. Opt. Soc. Am. [A]*, *17*(3), 538–544.

Deeb, S. S., Lindsey, D. T., Hibiya, Y., Sanocki, E., Winderickx, J., Teller, D. Y., et al. (1992). Genotype-phenotype relationships in human red/green color-vision defects: Molecular and psychophysical studies. *Am. J. Hum. Genet.*, *51*, 687–700.

DeVries, H. L. (1946). Luminosity curve of trichromats. *Nature (Lond.)*, *157*, 736–737.

Dobkins, K. R., Thiele, A., & Albright, A. D. (2000). Comparison of red-green equiluminance points in humans and macaques: Evidence for different L:M cone ratios between species. *J. Opt. Soc. Am. [A]*, *17*, 545–556.

Falls, H. F., Wolter, R., & Alpern, M. (1965). Typical total monochromasy: A histological and psychophysical study. *Arch. Ophthalmol.*, *74*, 610–616.

Galezowski, X. (1868). *Du diagnostic des Maladies des Yeux par la Chromatoscopie rétinienne: Précéde d'une Etude sur les Lois physiques et physiologiques des Couleurs*. Paris: J.B. Baillière et Fils.

Gao, H., & Hollyfield, J. G. (1992). Aging of the human retina: Differential loss of neurons and retinal pigment epithelial cells. *Invest. Ophthalmol. Vis. Sci.*, *33*(1), 1–17.

Glickstein, M., & Heath, G. G. (1975). Receptors in the monochromat eye. *Vis. Res.*, *15*, 633–636.

Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina: II. Hue and saturation. *J. Opt. Soc. Am.*, *67*, 202–207.

Green, D. G. (1970). Regional variations in the visual acuity for interference fringes on the retina. *J. Physiol. Lond.*, *207*, 351–356.

Gunther, K. L., Neitz, J., & Neitz, M. (2006). A novel mutation in the short-wavelength-sensitive cone pigment gene associated with a tritan color vision defect. *Vis. Neurosci.*, *23*, 403–409.

Hagstrom, S. A., Neitz, J., & Neitz, M. (1998). Variations in cone populations for red-green color vision examined by analysis of mRNA. *NeuroReport*, *9*(9), 1963–1967.

Harrison, R., Hoefnagel, D., & Hayward, J. N. (1960). Congenital total color blindness, a clinicopathological report. *Arch. Ophthalmol.*, *64*, 685–692.

Hayashi, T., Motulsky, A. G., & Deeb, S. S. (1999). Position of a 'green-red' hybrid gene in the visual pigment array determines colour-vision phenotype. *Nat. Genet.*, *22*(May), 90–93.

Hess, R. F., Sharpe, L. T., & Nordby, K. (Eds.). (1990). *Night vision: Basic, clinical and applied aspects*. Cambridge, UK: Cambridge University Press.

Hofer, H., Carroll, J., Neitz, J., Neitz, M., & Williams, D. R. (2005). Organization of the human trichromatic cone mosaic. *J. Neurosci.*, *25*(42), 9669–9679.

Jacobs, G. H., & Neitz, J. (1993). Electrophysiological estimates of individual variation in the L/M cone ratio. In B. Drum (Ed.), *Colour vision deficiencies XI*. Dordrecht: Kluwer.

Jagla, W. M., Jägle, H., Hayashi, T., Sharpe, L. T., & Deeb, S. S. (2002). The molecular basis of dichromatic color vision in males with multiple red and green visual pigment genes. *Hum. Mol. Genet.*, *11*, 23–32.

Jones, B. W., & Marc, R. E. (2005). Retinal remodeling during retinal degeneration. *Exp. Eye Res.*, *81*, 123–137.

Kalmus, H. (1955). The familial distribution of congenital tritanopia, with some remarks on some similar conditions. *Ann. Hum. Genet.*, *20*, 39–56.

Kazmi, M. A., Sakmar, T. P., & Ostrer, H. (1997). Mutation of a conserved cysteine in the X-linked cone opsins causes color vision deficiencies by disrupting protein folding and stability. *Invest. Ophthalmol. Vis. Sci.*, *38*(6), 1074–1081.

Knau, H., Jägle, H., & Sharpe, L. T. (2001). L/M cone ratios as a function of retinal eccentricity. *Color Res. Appl.*, *26*, S128-S132.

Komaromy, A. M., Alexander, J. J., Chiodo, V. A., Hauswirth, W. W., Acland, G. M., & Aguirre, G. D. (2007). Cone-directed gene therapy with rAAV leads to restoration of cone function in a canine model of achromatopsia. *Invest. Ophthalmol. Vis. Sci.*, *48*, E-Abstract 4614.

König, A. (1894). Uber den menschlichen Sehpurpur und seine Bedeutung fur das Sehen. *S.B. Akad. Wiss. Berlin*, 577–598.

Kremers, J., Scholl, H. P. N., Knau, H., Berendschot, T. T. J. M., Usui, T., & Sharpe, L. T. (2000). L/M cone ratios in human trichromats assessed by psychophysics, electroretinography, and retinal densitometry. *J. Opt. Soc. Am. [A]*, *17*, 517–526.

Larsen, H. (1921). Demonstration mikroskopischer Präparate von einem monochromatischen Auge. *Klin. Monatsbl. Augenheilkd.*, *67*, 301–302.

Makous, W., Carroll, J., Wolfing, J. I., Lin, J., Christie, N., & Williams, D. R. (2006). Retinal microscotomas revealed with adaptive-optics microflashes. *Invest. Ophthalmol. Vis. Sci.*, *47*(9), 4160–4167.

Marcos, S., & Navarro, R. (1997). Determination of the foveal cone spacing by ocular speckle interferometry: Limiting factors and acuity predictions. *J. Opt. Soc. Am. [A]*, *14*(4), 731–740.

Mariani, A. P. (1984). Bipolar cells in monkey retina selective for the cones likely to be blue-sensitive. *Nature*, *308*, 184–186.

McLellan, J. S., Marcos, S., Prieto, P. M., & Burns, S. A. (2002). Imperfect optics may be the eye's defence against chromatic blur. *Nature*, *417*(6885), 174–176.

Miyahara, E., Pokorny, J., Smith, V. C., Baron, R., & Baron, E. (1998). Color vision in two observers with highly biased LWS/MWS cone ratios. *Vis. Res.*, *38*(4), 601–612.

Miyake, Y., Yagasaki, K., & Ichikawa, H. (1985). Differential diagnosis of congenital tritanopia and dominantly inherited juvenile optic atrophy. *Arch. Ophthalmol.*, *103*, 1496–1501.

Mollon, J. D., & Bowmaker, J. K. (1992). The spatial arrangement of cones in the primate fovea. *Nature*, *360*, 677–679.

Mullen, K. T. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *J. Physiol. Lond.*, *359*, 381–400.

Mullen, K. T. (1991). Colour vision as a post-receptoral specialization of the central visual field. *Vis. Res.*, *31*, 119–130.

Nathans, J., Davenport, C. M., Maumenee, I. H., Lewis, R. A., Hejtmancik, J. F., Litt, M., et al. (1989). Molecular genetics of human blue cone monochromacy. *Science*, *245*, 831–838.

Nathans, J., Maumenee, I. A., Zrenner, E., Sadowski, B., Sharpe, L. T., Lewis, R. A., et al. (1993). Genetic heterogeneity among blue-cone monochromats. *Am. J. Hum. Genet.*, *53*, 987–1000.

Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B., & Hogness, D. S. (1986). Molecular genetics of inherited variation in human color vision. *Science*, *232*, 203–210.

Nathans, J., Thomas, D., & Hogness, D. S. (1986). Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science*, *232*, 193–202.

Neitz, J., Carroll, J., Yamauchi, Y., Neitz, M., & Williams, D. R. (2002). Color perception is mediated by a plastic neural mechanism that is adjustable in adults. *Neuron*, *35*(4), 783–792.

Neitz, M., Balding, S. D., McMahon, C., Sjoberg, S. A., & Neitz, J. (2006). Topography of long- and middle-wavelength sensitive cone opsin gene expression in human and Old World monkey retina. *Vis. Neurosci.*, *23*, 379–385.

Neitz, M., Carroll, J., Renner, A., Knau, H., Werner, J. S., & Neitz, J. (2004). Variety of genotypes in males diagnosed as dichromatic on a conventional clinical anomaloscope. *Vis. Neurosci.*, *21*, 205–216.

Nishiguchi, K. M., Sandberg, M. A., Gorji, N., Berson, E. L., & Dryja, T. P. (2005). Cone cGMP-gated channel mutations and clinical findings in patients with achromatopsia, macular degeneration, and other hereditary cone diseases. *Hum. Mutat.*, *25*, 248–258.

Packer, O., & Williams, D. R. (2003). Light, the retinal image, and photoreceptors. In S. K. Shevell (Ed.), *The science of color* (2nd ed., pp. 41–102). New York: OSA & Elsevier Science.

Packer, O. S., Williams, D. R., & Bensinger, D. G. (1996). Photopigment transmittance imaging of the primate photoreceptor mosaic. *J. Neurosci.*, *16*(7), 2251–2260.

Poirson, A. B., & Wandell, B. A. (1993). Appearance of colored patterns: Pattern-color separability. *J. Opt. Soc. Am. [A]*, *10*(12), 2458–2470.

Pokorny, J., Smith, V. C., & Verriest, G. (1979). Congenital color defects. In J. Pokorny, V. C. Smith, G. Verriest, & A. J. L. G. Pinckers (Eds.), *Congenital and acquired color vision defects* (pp. 183–241). New York: Grune & Stratton.

Pokorny, J., Smith, V. C., & Went, L. N. (1981). Color matching in autosomal dominant tritan defect. *J. Opt. Soc. Am.*, *71*, 1327–1334.

Pokorny, J., Smith, V. C., & Wesner, M. F. (1991). Variability in cone populations and implications. In A. Valberg & B. B. Lee

(Eds.), *From pigments to perception: Advances in understanding visual processes* (pp. 23–34). New York: Plenum Press.

Pum, D., Ahnelt, P. K., & Grasl, M. (1990). Iso-orientation areas in the foveal cone mosaic. *Vis. Neurosci.*, *5*, 511–523.

Putnam, N. M., Hofer, H. J., Doble, N., Chen, L., Carroll, J., & Williams, D. R. (2005). The locus of fixation and the foveal cone mosaic. *J. Vis.*, *5*(7), 632–639.

Richards, J. E., Scott, K. M., & Sieving, P. A. (1995). Disruption of conserved rhodopsin disulfide bond by Cys187Tyr mutation causes early and severe autosomal dominant retinitis pigmentosa. *Ophthalmology*, *102*(4), 669–677.

Rodieck, R. W. (1991). The density recovery profile: A method for the analysis of points in the plane applicable to retinal studies. *Vis. Neurosci.*, *6*, 95–111.

Roorda, A., Metha, A. B., Lennie, P., & Williams, D. R. (2001). Packing arrangement of the three cone classes in primate retina. *Vis. Res.*, *41*, 1291–1306.

Roorda, A., & Williams, D. R. (1999). The arrangement of the three cone classes in the living human eye. *Nature (Lond.)*, *397*, 520–522.

Rushton, W. A. H., & Baker, H. D. (1964). Red/green sensitivity in normal vision. *Vis. Res.*, *4*, 75–85.

Sakmar, T. P. (2002). Structure of rhodopsin and the superfamily of seven-helical receptors: The same and not the same. *Curr. Opin. Cell Biol.*, *14*(2), 189–195.

Samy, C. N., & Hirsch, J. (1989). Comparison of human and monkey retinal photoreceptor sampling mosaics. *Vis. Neurosci.*, *3*, 281–285.

Sekiguchi, N., Williams, D. R., & Brainard, D. H. (1993a). Aberration-free measurements of the visibility of isoluminant gratings. *J. Opt. Soc. Am. [A]*, *10*, 2105–2117.

Sekiguchi, N., Williams, D. R., & Brainard, D. H. (1993b). Efficiency in detection of isoluminant and isochromatic interference fringes. *J. Opt. Soc. Am. [A]*, *10*, 2118–2133.

Szel, A., Diamanstein, T., & Rohlich, P. (1988). Identification of blue-sensitive cones in the mammalian retina by antivisual pigment antibody. *J. Comp. Neurol.*, *273*, 593–602.

Ueyama, H., Li, Y.-H., Fu, G.-L., Lertrit, P., Atchaneeyasakul, L., Oda, S., et al. (2003). An A-71C substitution in a green gene at the second position in the red/green visual-pigment gene array is associated with deutan color-vision deficiency. *Proc. Natl. Acad. Sci. USA*, *100*(6), 3357–3362.

Weitz, C. J., Miyake, Y., Shinzato, K., Montag, E., Zrenner, E., Went, L. N., et al. (1992). Human tritanopia associated with two amino acid substitutions in the blue sensitive opsin. *Am. J. Hum. Genet.*, *50*, 498–507.

Weitz, C. J., Went, L. N., & Nathans, J. (1992). Human tritanopia associated with a third amino acid substitution in the blue sensitive visual pigment. *Am. J. Hum. Genet.*, *51*, 444–446.

Went, L. N., & Pronk, N. (1985). The genetics of tritan disturbances. *Hum. Genet.*, *69*, 255–262.

Williams, D. R., & Collier, R. J. (1983). Consequences of spatial sampling by a human photoreceptor mosaic. *Science*, *221*, 385–387.

Williams, D. R., MacLeod, D. I. A., & Hayhoe, M. (1981a). Foveal tritanopia. *Vis. Res.*, *21*(9), 1341–1356.

Williams, D. R., MacLeod, D. I. A., & Hayhoe, M. M. (1981b). Punctate sensitivity of the blue-sensitive mechanism. *Vis. Res.*, *21*, 1357–1375.

Williams, D. R., Sekiguchi, N., Haake, W., Brainard, D. H., & Packer, O. (1991). The cost of trichromacy for spatial vision. In B. B. Lee & A. Valberg (Eds.), *From pigments to perception: Advances in understanding visual processes* (pp. 11–22). New York: Plenum Press.

Williams, R. W. (1991). The human retina has a cone-enriched rim. *Vis. Neurosci.*, *6*(4), 403–406.

Willmer, E. N., & Wright, W. D. (1945). Colour sensitivity of the fovea centralis. *Nature*, *156*, 119–121.

Winderickx, J., Sanocki, E., Lindsey, D. T., Teller, D. Y., Motulsky, A. G., & Deeb, S. S. (1992). Defective colour vision associated with a missense mutation in the human green visual pigment gene. *Nat. Genet.*, *1*, 251–256.

Wright, W. D. (1952). The characteristics of tritanopia. *J. Opt. Soc. Am.*, *42*, 509–521.

Yellott, J. I., Jr., Wandell, B., & Cornsweet, T. (1984). The beginnings of visual perception: The retinal image and its initial coding. In I. Darian-Smith (Ed.), *Handbook of physiology, Section 1: The nervous system III* (pp. 257–316). Bethesda, MD: American Physiological Society.

27 Bayesian Approaches to Color Vision

DAVID H. BRAINARD

ABSTRACT Visual perception is difficult because image formation and sensory transduction lose information about the physical scene: Many different scenes lead to the same image data. Understanding how the brain copes with this information loss, so that our percepts provide a useful representation of the world around us, is a central problem in cognitive neuroscience. In the case of color vision, the nature of the information loss is well understood. First, the light reflected to the eye confounds illuminant properties with those of objects. Second, spectral and spatial sampling by the cone photoreceptors further reduces the available information. To provide a stable representation of object color, the brain must compensate by combining the directly available information with assumptions about which scene configurations are likely to occur. This chapter reviews how Bayesian decision theory can model how this happens and discusses two Bayesian models that have been effective in accounting for color appearance.

Visual perception is difficult. One pervasive reason for this difficulty is that image formation and sensory transduction lose information about the physical scene, so many different scenes could have caused the same image data. Color vision presents an opportunity to understand how the brain copes with this information loss, because our understanding of the information loss and the scene parameters of perceptual interest is well developed. In this sense, color provides a model system for developing and testing theories that may have more general applicability. Of course, color perception is an important aspect of our perceptual experience, and understanding how it arises is also interesting in its own right. This chapter provides an introduction to Bayesian modeling of human color vision and reviews two lines of work where the approach has been fruitful.

Fundamentals of color vision

The visual system assigns a color to essentially all viewed objects. The information available about an object's color is carried by the spectrum of the light reflected from it, as illustrated by figure 27.1*A*. This spectrum, which we refer to as the *color signal*, is specified by its power at each wavelength.

DAVID H. BRAINARD Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania

The color signal is given as the wavelength-by-wavelength product of the illuminant power and the object's surface reflectance function, where the latter specifies the fraction of incident light reflected from the object. Thus the color signal confounds object properties with those of the illuminant. To provide a representation of object reflectance that is stable across changes of illuminant, the visual system must process the color signal to separate the physical effects of illuminant and object surface.

The postreceptoral visual system does not have direct access to the color signal. Rather, this spectrum is encoded by the joint responses of the retinal mosaic of cone photoreceptors. There are three classes of cones, each characterized by a distinct spectral sensitivity (figure 27.1*B*). These are often referred to as the L, M, and S cones.

Each individual cone codes information about light as a scalar quantity, the rate at which its photopigment is isomerized. This rate confounds the overall intensity of the color signal with its relative spectrum. Thus two physically distinct color signals can produce the same isomerization rate in all three classes of cones (figure 27.1*C*). Moreover, there is at most one cone at each retinal location. To obtain even trichromatic information about the spectrum of the color signal, the visual system must combine information from cones at different retinal locations; sampling of the image by the retinal mosaic confounds spatial and chromatic image structure.

This brief review illustrates a series of stages in which information about object spectral properties is lost: The color signal confounds object reflectance with the spectral power distribution of the illuminant; the retina as a whole contains only three classes of univariate cones, so the color signal's full spectrum is represented by at most three numbers; and at each retinal location, there is only one cone. Each of these stages of information loss produces ambiguity about the scene being viewed.

How does the visual system resolve ambiguity to extract a perceptually useful representation of object color? Since the cone responses do not completely determine the reflectance properties of the object, some additional constraints must be imposed. Here, we apply Bayesian analysis as a framework to express these constraints and develop models of color perception.

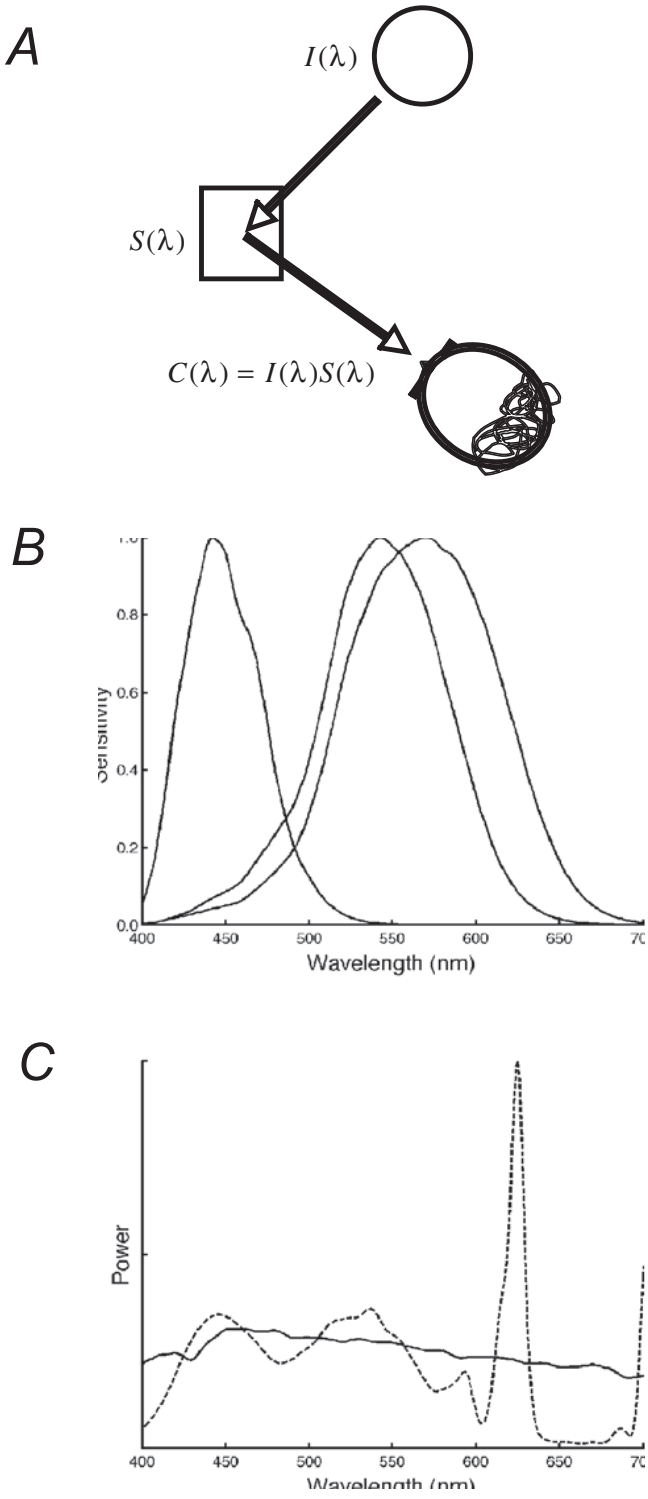


FIGURE 27.1 Information loss in color vision. (*A*) Color vision begins when light reflects from an object to the eye and is imaged on the retina. The color signal reaching each location of the retina is characterized by its spectrum, which is the amount of power at each wavelength. (*B*) The spectral sensitivities of the L, M, and S cones. The isomerization rate of photopigment is determined by multiplying the incident power of the color signal at each wavelength by the photopigment's spectral sensitivity and summing over wavelength. (*C*) Two physically distinct spectra that produce the same isomerization rates in the human L, M, and S cones simultaneously.

Bayesian principles

BASIC IDEAS Bayesian statistics provides a general formulation that allows image data to be combined with prior assumptions to provide a reasonable estimate of the physical scene configuration. Both the information provided by the data and the prior assumptions are expressed in the common language of probability distributions. These two types of information are then combined via Bayes' rule to produce a *posterior* distribution that expresses what is known about the scene. A specific estimate of the scene configuration can then be extracted from the posterior, for example, by taking its mean.

An example serves to illustrate the key ideas. Imagine a toy universe containing only one wavelength of light and one spatial location. An illuminant impinging on a surface is specified by an intensity i, and the object surface reflectance is specified by a single number s. We refer to these as the *scene parameters*, as fixing their values specifies the physical scene. The color signal here reduces to a single number, $c = is$. If an eye with a single photoreceptor images the scene, we can model its response r by the equation $r = c + n$, where n is additive noise. We then ask, "What do the *image data* r tell us about the values of the scene parameters?" Within the framework of Bayesian analysis, this is given by the *likelihood*, written $p(r|\, i, s)$. For any scene parameters, the likelihood tells us how probable any observed response r is. Figure 27.2*A* illustrates the likelihood for our example, for the case $r = 1$. When the image data are held fixed, the likelihood is a function of the scene parameters.

Two features of the likelihood are worth note. First, some pairs (i, s) lead to higher likelihood than others. This means that the image data provide some information about the scene parameters. Second, the likelihood function has a ridge of equal values along the hyperbola $is = 1$. The fact that the likelihood is equal along this ridge indicates that the image data provide incomplete information; there are multiple scene configurations that the image data do not distinguish.

Because the image data do not uniquely determine the scene parameters, some other principle must be invoked to resolve the ambiguity. The prescription provided by the Bayesian approach is to specify the statistical properties of the scene parameters. In our toy universe, for example, it might be that not all illuminant intensities occur with equal probability. This fact can be expressed as a *prior* probability distribution over the scene parameters (figure 27.2*B*). Here, the prior has a ridge parallel to the s dimension, indicating that all values of s are equally likely. But along the i dimen-

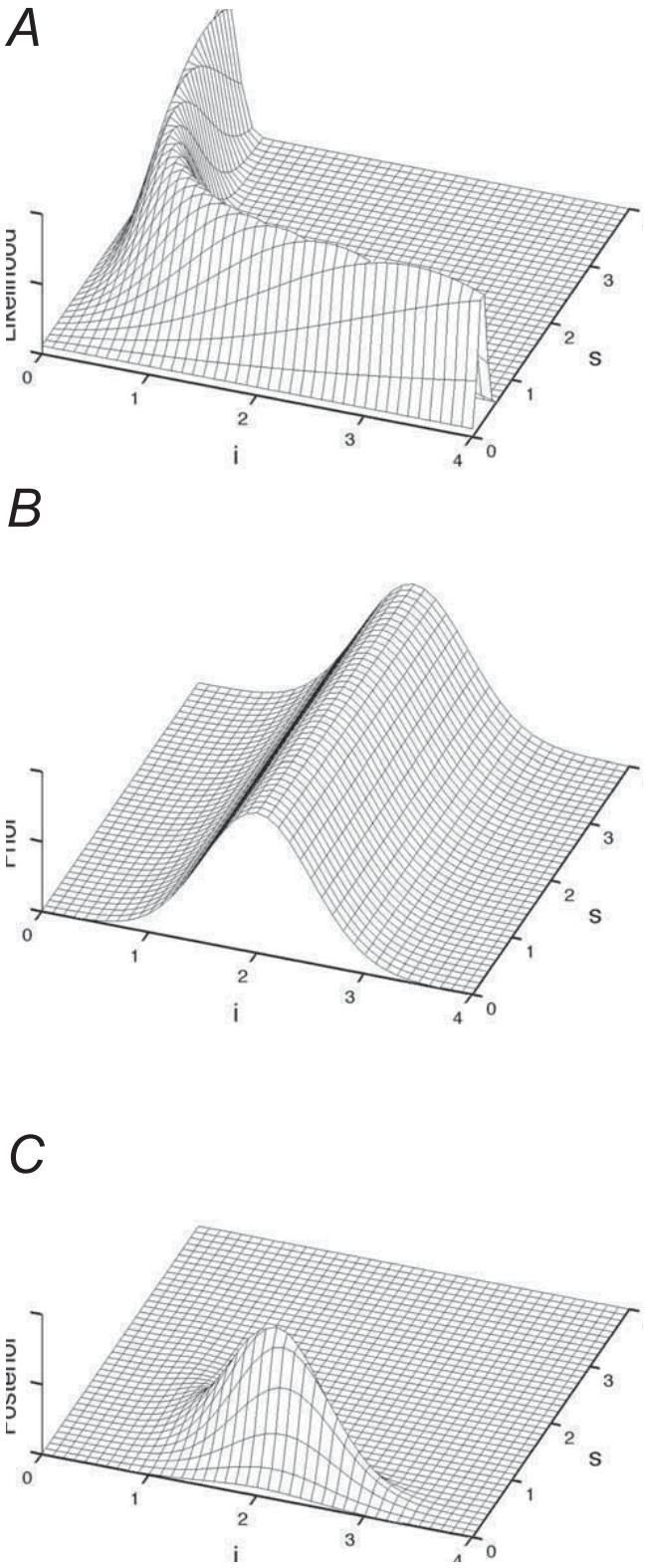


FIGURE 27.2 Simple example of Bayesian analysis. (*A*) Likelihood function for simple example. The likelihood function $p(r|i, s)$ is plotted as a function of (i, s) for the case in which $r = 1$. Values of i and s are assumed to be constrained between 0 and 4. (*B*) Prior for simple example. The prior probability distribution $p(i, s)$ is plotted, showing a prior where illuminant values are more likely near $i = 2$ but that is silent about surface reflectance s. (*C*) The posterior is given by the pointwise product of the likelihood and the prior and then is normalized. The resulting $p(i|s, r)$ is shown for the likelihood in panel *A* and the prior in panel *B*. For additional discussion of this example, see Brainard and Freeman (1997).

sion, there is a concentration of illuminant probability in the vicinity of $i = 2$. The prior is a function of the same scene parameters as the likelihood.

Bayes' rule says that to combine the prior and likelihood into the posterior, all that is necessary is to multiply the two functions point by point and normalize the result so that the total probability is unity (Gelman, Carlin, Stern, & Rubin, 2004). The result of this operation for our example is illustrated by figure 27.2*C*. The use of multiplication makes intuitive sense: the posterior has high values when both the prior and the likelihood are high and goes to zero if either the prior or the likelihood is zero. The posterior is often written in the form $p(i, s|r)$. In the case shown, the posterior has an unambiguous peak, even though both the likelihood and the prior are individually ambiguous about the scene parameters. Often, the scene parameters are estimated by the maximum or as the mean of the posterior; other methods are also available (Blackwell & Girschick, 1954; Brainard & Freeman, 1997; Maloney, 2002).

BAYESIAN MODELS Bayesian analysis provides a framework for generating models that may be applied to specific perceptual phenomena. The task of the modeler is to express the content of interest as a likelihood and prior and then to link the resulting estimate of the scene parameters to perception. The framework is useful to the extent that it consistently generates models that describe, predict, and clarify empirical data.

The first part of Bayesian modeling is to specify the likelihood, which amounts to understanding the process by which scene parameters determine image data. For vision, the likelihood is in essence a description of how light flows through a scene to produce the retinal image, how the retinal image is sampled by the photoreceptor mosaic, and how precise the photoreceptor responses are. In the case of color, these factors are well understood (Wyszecki & Stiles, 1982; Kaiser & Boynton, 1996; Wandell, 1995), which means that generating the likelihood portion of a Bayesian model is straightforward.

The second part of Bayesian modeling is to specify the prior. In cases in which the likelihood is well constrained,

the prior carries the critical content of the model. One way to obtain a prior is to examine statistical regularities in the natural environment and to choose a prior that captures these regularities. The idea is that evolution and development have optimized visual processing for the environment in which we operate and that statistical regularities in the environment are likely to be deeply embedded in visual processing (see Helmholtz, 1910; Attneave, 1954; Shepard, 1987, 1992; Adelson & Pentland, 1996; Mamassian, Landy, & Maloney, 2002; Weiss, Simoncelli, & Adelson, 2002; Geisler & Kersten, 2002; Maloney, 2002; see also Knill & Richards, 1996; Rao, Olshausen, & Lewicki, 2002).

Often, not enough is known about natural scene statistics to completely determine a prior. In this case, one strategy is to choose a parametric form for the prior that is broadly consistent with available measurements and then to use psychophysical data to constrain these parameters. We (Brainard et al., 2006; Brainard, Williams, & Hofer, 2008) and others (e.g., Mamassian & Landy, 2001; Stocker & Simoncelli, 2006) have found that this hybrid approach leads to effective models and allows good progress to be made with the Bayesian approach even as we seek improved methods for measuring and specifying priors.

We defer to the discussion a number of important issues related to the selection of priors in Bayesian models of perception. Such discussion is more cogent in the context of the specific models reviewed in this chapter.

The final step in developing a Bayesian model of perception is to link the output of the Bayesian calculation, which is typically an estimate of scene parameters extracted from the posterior, to a measurement of human performance. Often, the linking hypothesis involves the assumption that that goal of perception is to provide an explicit representation of one of the scene parameters. For example, color appearance is often taken to be a perceptual correlate of object surface reflectance. Given this idea, a natural linking hypothesis for object color appearance is that two objects, viewed in different scenes, will have the same color appearance when a Bayesian estimation algorithm returns the same estimate of surface reflectance for each (Brainard, Kraft, & Longère, 2003).

Color constancy

In daily life, it is common to refer to objects as having a color appearance: "the red apple," "the blue house," "the green car." Although this is effortless, it is also remarkable. The color signal that is reflected from any object varies with the illumination, so a stable percept must involve postreceptoral processing. The stability of object color appearance is called *color constancy*. Both color constancy and its close cousin, lightness constancy, have long been the target of experimental investigation (e.g., Helson & Jeffers, 1940; Burnham, Evans, & Newhall, 1957; McCann, McKee, & Taylor, 1976; Arend & Reeves, 1986; see Katz, 1935; Brainard, 2004; Gilchrist, 2006).

A few empirical generalizations are easy to state. First, the data confirm the introspective conclusion that human vision often exhibits excellent color constancy: Object color appearance changes less than would be predicted by the corresponding change in reflected light (Arend & Reeves, 1986; Brainard, Brunt, & Speigle, 1997; Brainard, 1998). Second, color constancy is not perfect: Object color appearance does change somewhat with the illuminant. Third, factors other than the illuminant affect object color appearance. A particularly important factor is the surface reflectances of other nearby objects in the scene (Kraft & Brainard, 1999; Gilchrist, 2006).

Given that the above general facts are well established, it seems clear that our goal should not be to generate further broad confirmations. Rather, we should seek models that accurately predict the color appearance of any object when it is viewed in an arbitrary scene. This is not an easy task, as a combinatorial explosion prevents enumeration and direct study of all possible scenes. It is therefore critical to build models that not only fit extant data but also embody principles that make it likely that they will generalize well (see Wyszecki & Stiles, 1982, pp. 584–586; see also Krantz, 1968; Brainard & Wandell, 1992).

Different approaches to developing models of color appearance may be distinguished by the nature of the core principles they embody, and how these might enable generalization. Mechanistic models are constrained by abstracted properties of neurons in the visual pathways (Stiles, 1967). Examples of such abstractions include chromatic adaptation (von Kries, 1902) and contrast coding (Wallach, 1948; Land, 1986). These models will generalize well if the mechanisms that are revealed by simple stimuli continue to operate unaltered for more complex images.

A second modeling approach tackles generalization directly and involves explicitly formulating and testing rules of combination. An example would be additive prediction of the action of the superposition of two scenes on an object's color appearance from measurements of the effect of each scene alone (Brainard & Wandell, 1992). A complement to this approach is to identify grouping and segmentation principles that allow the decomposition of complex scenes into separate and simpler regions and then to study effects of context within such regions and to consider how they interact (Gilchrist et al., 1999; Adelson, 1999).

A third approach is *computational* (Marr, 1982; Barrow & Tenenbaum, 1978). Here, the guiding hypothesis is that perceptual representations may be understood as biological approximations to well-defined information processing tasks. The use of Bayesian algorithms to model perception is an example of this approach. It will generalize well if the under-

lying hypothesis is correct: As one moves to richer and richer stimulus configurations, what must be updated is the modeled likelihood and prior, but no new idea is required.

A Bayesian model of human color constancy

Here, we outline a Bayesian model of color appearance. The first step is to define the likelihood function. This is based on the following imaging model (Brainard & Freeman, 1997; Brainard et al., 2006). A collection of flat matte objects is lit by a single spatially uniform and diffuse illuminant. The jth object is characterized by its surface reflectance $s_j(\lambda)$, which specifies the fraction of incident light reflected at each wavelength λ. The illuminant is characterized by its spectral power distribution $i(\lambda)$, which provides light power at each wavelength. The color signal reflected from the jth object to the eye at each wavelength is simply $c_j(\lambda) = i(\lambda)\, s_j(\lambda)$. For this example, we assume that the spatial scale of each object is large enough that we may neglect the fine structure of the interleaved retinal mosaic, so the information obtained from the color signal reflected from each object is the isomerization rates of the L, M, and S cones. These are easily computed from $c_j(\lambda)$ and the known spectral sensitivities of each cone class.

The isomerization rates arising from a given $c_j(\lambda)$ will be perturbed by noise, which can be modeled as additive and normally distributed. Given the surface reflectances $s_j(\lambda)$ of N objects in the scene and the illuminant spectral power distribution $i(\lambda)$, the likelihood may then be computed as a multivariate normal distribution with dimension $3N$, whose mean is given by the mean isomerization of the L, M, and S cones for each object and whose covariance matrix represents the additive noise. As with the simple example of figure 27.2, this likelihood function leads to an underdetermined estimation problem.

To specify a prior distribution for color constancy, we start with separate distributions over surface reflectance functions and illuminant spectral power distributions. The same general form is used for both and is illustrated for surfaces in figure 27.3. Following a number authors (Cohen, 1964; Buchsbaum, 1980; Maloney & Wandell, 1986; see also Maloney, 1986; Jaaskelainen, Parkkinen, & Toyooka, 1990), we first assume that each surface reflectance function $s_j(\lambda)$ may be approximated within a three-dimensional linear model. This means that we can write $s_j(\lambda) = w^1_j\, s^1(\lambda) + w^2_j\, s^2(\lambda) + w^3_j\, s^3(\lambda)$, where the three *basis functions* $s^i(\lambda)$ are held fixed (see figures 27.3*A* and 27.3*B*). The basis functions can be determined by applying principle components analysis to a large set of measured surface reflectance functions.

Within the three-dimensional model, each surface is specified by its triplet of weights w^1_j, w^2_j, w^3_j. To induce a prior distribution over surfaces, we measure the three-dimensional histogram of weights corresponding to an ensemble of surfaces and fit this with a multivariate normal distribution (figure 27.3*C*) (Brainard & Freeman, 1997). Doing so allows us to compute the prior probability of any surface reflectance $s_j(\lambda)$. The same approach can be used to determine prior probabilities for illuminants $i(\lambda)$.

Delahunt and Brainard (2004b) conducted experiments in which observers were asked to judge the appearance of a test patch embedded in the simulated images of 17 scenes. Images of four of the scenes are shown in figure 27.4*A*, the location of the test patch being indicated by the black rectangle in each image. Observers adjusted the chromaticity of the test patch until it appeared achromatic. Psychophysical measurements of this sort establish the chromaticity of the *achromatic locus* at the test patch location within each image, and this provides an excellent first-order characterization of how scene context affects the color of any object (Speigle & Brainard, 1999). Across the set of scenes, both the illuminant and the surface reflectance of the background surface were varied. Figure 27.4*B* plots the chromaticity of the illuminant for each example scene (large open circles) along with the achromatic loci (large solid circles). For three of the scenes, the chromaticity of the achromatic locus is near to that of the illuminant. These are the scenes in which only the illuminant varied, and this pattern indicates good color constancy (see Delahunt & Brainard, 2004b; Brainard et al., 2006). For the fourth scene, in which the background reflectance was also manipulated, the achromatic locus is far from the illuminant chromaticity; constancy was poor under this manipulation.

Brainard and colleagues (2006) used the general form of likelihood and prior described above and implemented a variant of Brainard and Freeman's (1997) Bayesian algorithm that estimates the chromaticity of the scene illuminant from image data. They used a prior for surfaces that captured the structure of a large set of measured reflectances (as in figure 27.3) and explored a parametric set of illuminant priors, all based on the CIE linear model for daylight. Brainard and colleagues (2006) then used the Bayesian algorithm to estimate the illuminant chromaticities for the same 17 scenes studied by Delahunt and Brainard (2004b). To link the illuminant estimates to the measured achromatic loci, they interpreted the achromatic loci as representing the chromaticity of light reflected from a fixed achromatic surface. This allowed mapping of estimated illuminant chromaticities to predictions of the achromatic loci (Brainard et al., 2006).

Figure 27.4*B* shows the predictions derived in this fashion (small open circles). The prediction for the leftmost image is guaranteed to be correct because this datum was used to infer the reflectance of the achromatic surface. The predictions for the other three scenes, and indeed for the entire collection of 17 images (see Brainard et al., 2006), are good.

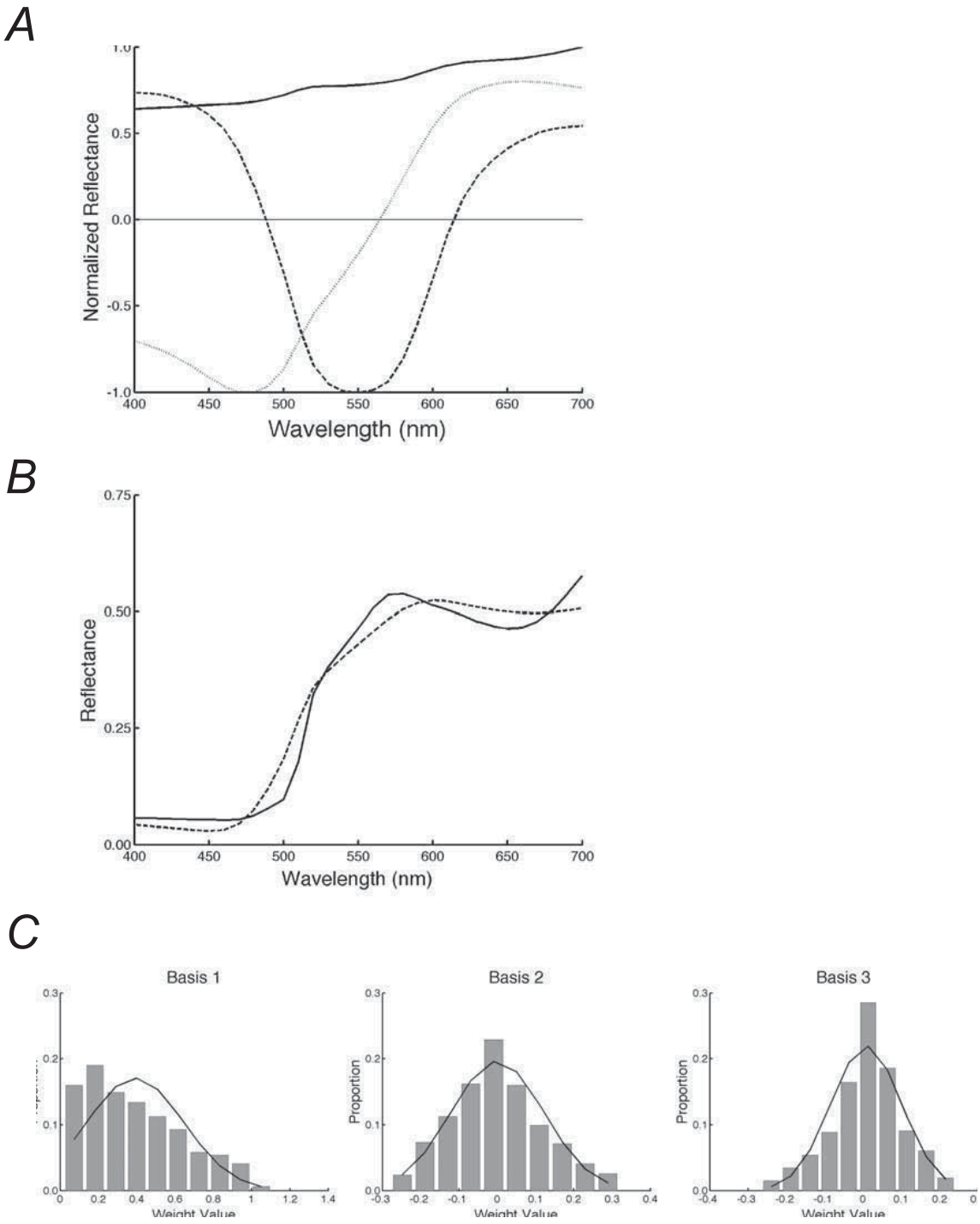


FIGURE 27.3 Prior over surface reflectance. (*A*) Three basis functions of a linear model for surfaces. The basis functions were obtained through analysis of the principle components of a collection of 462 measured surface reflectance functions (Newhall, Nickerson, & Judd, 1943; Nickerson, 1957). (*B*) A measured surface reflectance function (solid curve) from the same data set used to generate the linear model, and its approximation (dashed curve) within the model shown in panel *A*. (*C*) Distribution of model weights for the 462 surface reflectance functions in the data set used to generate the linear model. Each panel shows the histogram for one basis function. The solid curves are a normal approximation. (After Brainard & Freeman, 1997; see their figure 4.)

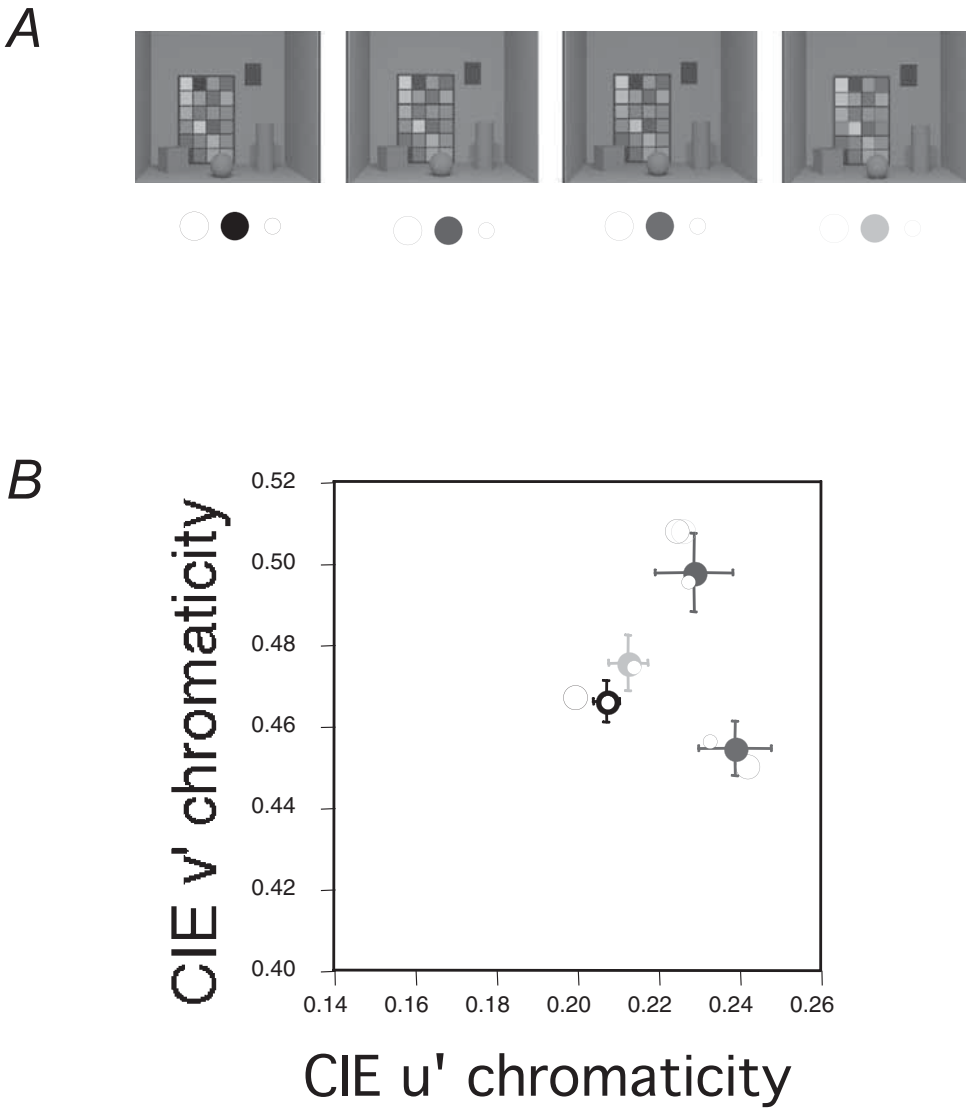


FIGURE 27.4 Color constancy performance. (*A*) Images of 4 of 17 simulated scenes used to compare human performance and a model derived from a Bayesian illuminant estimation algorithm. Each scene has the same spatial structure. In the first three images, from left to right, the illuminant varies. In the rightmost image, the illuminant is the same as that in the second image, but the background surface has been changed so that the light reflected from it matches that reflected from the background surface in the leftmost image. (Reproduced from Brainard et al., 2006, figure 1.) (*B*) Illuminants, achromatic loci, and model predictions plotted in the CIE $u'v'$ chromaticity diagram. This is a standard color representation that preserves information about the relative responses of the L, M, and S cones but not about intensity. Large open circles show the scene illuminants, with the color key as indicated beneath the images in panel *A*. Large solid circles show the achromatic loci measured by observers who adjusted a test patch at the location indicated by the black rectangle in each image. The small open circles show the model's predictions of the achromatic loci. (Reproduced from figure 7 of Brainard et al., 2006.) (See color plate 38.)

Of importance is that the agreement occurs both for cases in which the observers showed good constancy and cases in which constancy was poor.

Bayes and the cone mosaic

The treatment of color constancy above addresses the information loss caused by the interaction of surfaces and illuminants in the formation of the image and the reduction from a full spectral representation of the color signal to the trichromatic representation provided by the L, M, and S cones. It assumed, however, that the responses of all three cone classes were available at each spatial location. This is intuitively reasonable if the spatial scale of the image of various objects is large in comparison to the spacing between cones. Nonetheless, it would be more satisfying to have a theory of how information from cones is combined across space to provide a representation that is effectively trichromatic at such spatial

scales. To develop such a theory, we can again turn to a Bayesian analysis.

As with the color constancy case, developing a Bayesian model requires specifying the likelihood and prior. We begin by deciding what scene parameters need to be estimated. Here, we take these to be an *ideal image*, which we conceive of as the isomerization rates of the L, M, and S cones at each image location, before optical blur or retinal sampling. From the ideal image, we can compute the mean isomerization rate for each of the actual cones in the interleaved mosaic. The computation incorporates effects of blurring by the eye's optics as well as the location and type of each individual cone. We then use an approximation of additive normal noise to convert the mean isomerization to the likelihood function.

To develop an appropriate prior, we rely on a few well-established observations about natural images. First, within a cone class, images vary slowly over space. This is often expressed by the observation that the energy in natural images falls off as $1/f$, where f is spatial frequency (Field, 1987). Alternatively, the same fact may be described as a strong correlation between image values at nearby locations within each cone class (Pratt, 1978). Second, at any given location, there are high correlations between the isomerization rates of the L, M, and S cones (Burton & Moorehead, 1987; Ruderman, Cronin, & Chiao, 1998). This correlation occurs because typical color signals are spectrally broadband and vary slowly with wavelength, because the spectral sensitivities of cones are themselves broad, and because the spectral sensitivities of different cone classes overlap. We instantiate these strong spatial and spectral correlations in our prior by using a multivariate normal distribution whose covariance matrix is separable in space and color. This allows expression of the appropriate correlations (Brainard, 1994; Brainard et al., 2008). The use of a normal prior does not describe all of the regularities of natural images, but is strong enough to enable an effective algorithm.

Combining the likelihood and prior described above allows an estimate (Brainard, 1994; Brainard et al., 2008) of the ideal image from the isomerization rates of cones in the mosaic. Figure 27.5 illustrates estimator performance for a low spatial frequency stimulus. For such stimuli, the estimator returns near veridical estimates of the ideal image. This justifies the typical assumption that the fine structure of the interleaved mosaic may be neglected for most treatments of color vision.

It is possible to present stimuli that cause individuals with different cone mosaics to have distinguishable perceptions. Hofer, Singer, and Williams (2005) used adaptive optics to present very small flashed monochromatic spots to observers whose chromatic topography had also been mapped. Figure 27.6 summarizes the results of these experiments. The top row (*A*) shows mosaics for five individual observers. The

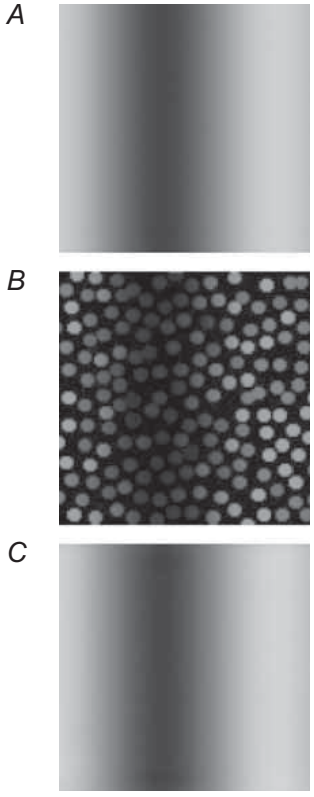
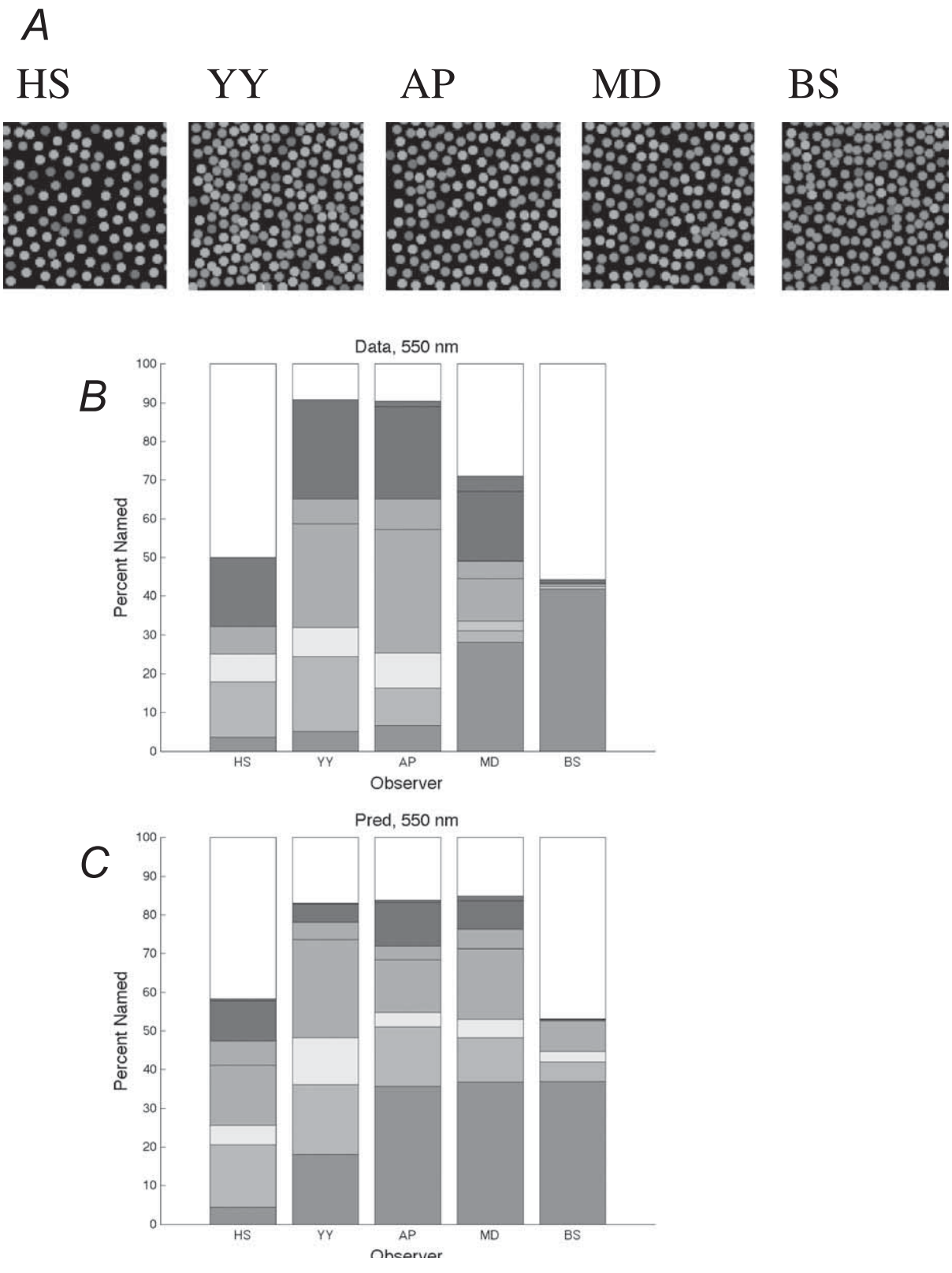


FIGURE 27.5 Trichromatic reconstruction. (*A*) Small patch of sinusoidal isochromatic grating. (*B*) The intensity of each colored spot represents the isomerization rate of a cone. The class of each cone is indicated by whether it is plotted in red (L), green (M), or blue (S). (*C*) The isochromatic grating as reconstructed by the Bayesian algorithm. The grating shown corresponds to a spatial frequency of 6 cycles per degree presented at about 1 degree of eccentricity for a human observer; the mosaic is of observer AP of Hofer and colleagues (2005). Brainard and colleagues (2008) provide additional reconstruction examples that show similarly veridical performance for low-spatial-frequency isoluminant gratings and for an additional mosaic. (See color plate 39.)

FIGURE 27.6 Small spot experiment. (*A*) Schematic of five individual observer cone mosaics. Red, green, and blue circles show locations of L, M, and S cones. Mosaics represent approximately 12 by 12 feet of visual angle at 1 degree of eccentricity. (Reproduced from Brainard et al., 2008, figure 2 (top panel).) (*B*) Data from Hofer and colleagues (2005) for 550-nm spots. Observers named each small spot that they saw and judged namable. The available names were red, orange, yellow, yellow-green, green, blue-green, purple, and white. For each observer, the histogram shows the proportion of each color named used, with the color code corresponding to the name. Note that the white region of the bars represents the proportion of white responses. Not all observers used all available names. (Reproduced from figure 11 of Brainard et al., 2008.) (*C*) Predictions from the Bayesian model, obtained as described in the text, for the experimental conditions corresponding to the data in *B*. (Reproduced from figure 11 of Brainard et al., 2008), which also shows data and predictions for 500-nm and 600-nm spots.) (See color plate 40.)


A
HS
YY
AP
MD
BS
B
Data, 550 nm
Percent Named
Observer
HS
YY
AP
MD
BS
C
Pred, 550 nm
Percent Named
Observer

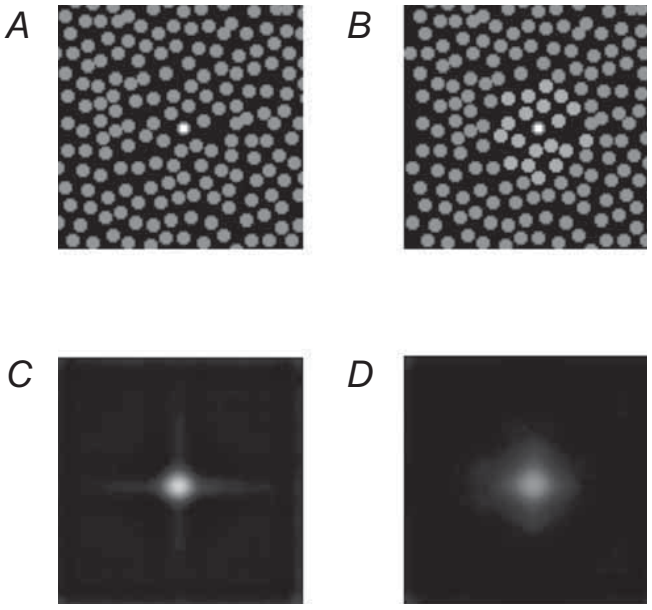


FIGURE 27.7 Small spot intuitions. (*A*) A mosaic consisting only of L cones. The white spot in the center indicates a single cone whose stimulation was simulated. (*B*) This mosaic is identical to the one shown in panel *A*, with the exception that the cones surrounding the central L cone have been changed to M cones. (*C*) Model output when the central cone in panel *A* is stimulated. The result is a bluish-white spot. As described by Brainard and colleagues (2008), a windowing procedure was applied to model output here and in panel *D* to reduce visible ringing in the reconstruction. (*D*) Model output when the central cone in panel *B* is stimulated. The result is a reddish spot. (Reproduced from figure 7 of Brainard et al., 2008.) (See color plate 41.)

second row (*B*) shows histograms of the aggregate naming behavior of each observer. There is large individual variation. Of particular note is that individuals with extreme L-to-M cone ratios tended to see more flashes as "white" than did individuals with more equal cone ratios.

We simulated the experiment of Hofer and colleagues (2005) and obtained reconstructed spots for each simulated flash. These were then mapped to color names through their chromaticities. Figure 27.6*C* shows the predicted naming histograms for each observer. The predictions capture the broad patterns of the data well. In particular, the fact that highly asymmetric mosaics lead to more flashes named "white" is a salient feature of the predictions, just as it is with the data.

Figure 27.7 provides intuition about the model's predictions. First consider the left-hand panels (*A* and *C*). These show a mosaic consisting only of L cones and the Bayesian estimate of the ideal image when only the central L cone is stimulated. A mosaic consisting of only L cones is completely monochromatic, and its responses provide no information about the relative spectral composition of the stimulus. When the Bayesian algorithm is applied to data provided by such a mosaic, it must rely on the prior to resolve the spectral ambiguity. The prior had a mean of CIE daylight illuminant D65, and the resulting estimate is correspondingly a bluish white.

The right-hand panels of figure 27.7 (*B* and *D*) show a second mosaic with the same spatial arrangement of cones. In this mosaic, however, the central L cone is surrounded by a set of M cones. We again simulated stimulating only the central L cone and obtained the algorithm's estimate. Here, the resulting spot is red. This occurs because the M cones near to the central L cone add information about the spectral composition of the ideal image that was not available in the all L cone mosaic. These M cones have isomerization rates of zero, indicating that there is no middle wavelength light at their locations. Because the prior includes a specification of strong spatial correlations within each color plane, the zero M cone responses say that had an M cone been present at the central location, it too would have had a small response. Putting this together with the large observed L cone response leads to an estimate of a spot with more power in the long wavelengths than in the middle wavelengths, and the result appears red.

Although the actual individual mosaics are more complex than the examples shown in figure 27.7, the same intuitions apply. In mosaics in which there are large regions of homogeneous cone types, the reconstructed spots will have to rely more on the prior mean and will tend to be "white" rather than a more saturated color. The detailed modeling results shown in figure 27.6 play this and related intuitions out in detail.

Discussion

The work reviewed here links color appearance to the performance of Bayesian algorithms. The first model (Brainard et al., 2006) accounts for both successes and failures of human color constancy. The second model (Brainard et al., 2008) explains how the visual system integrates information from individual cones in the retinal mosaic to provide a seamless trichromatic percept under normal viewing conditions. This model successfully predicts the color appearance of spots that are small enough to stimulate single cones.

The two models have important common features. First, the modeling begins with an analysis of information loss between the scene parameters and the responses of the cones. Second, the core of each model is the specification of a prior distribution over the scene parameters of interest. The prior acts to resolve the ambiguity about the scene parameters that remains after the image data have spoken. Given that the likelihood is well constrained, it is the prior that provides the content of each model. In both cases, the general parametric form of the prior was determined through an analysis of naturally occurring scenes. The particular

priors used to model the data were then tuned by comparison of model predictions to the data (see Brainard et al., 2006, 2008).

The Bayesian approach leads to parsimonious models that account for a performance across a wide range of conditions. Consider the color constancy example. Many studies of color constancy vary only the illumination, while the surfaces in the scene are held fixed (e.g., McCann, McKee, & Taylor, 1976; Arend & Reeves, 1986; Brainard & Wandell, 1992; Hansen, Walter, & Gegenfurtner, 2007). Under these conditions, many computational models can predict the generally good constancy that is observed (Buchsbaum, 1980; Land, 1986; Brainard & Freeman, 1997). When the surfaces in the scene are covaried with the illuminant, constancy is a more challenging computational problem that can differentiate between models (Maloney & Wandell, 1986; Brainard & Wandell, 1986; Brainard & Freeman, 1997; Kraft & Brainard, 1999). The model presented here accounts for performance when the illuminant is varied alone and when both illuminant and scene surfaces are covaried. This feature emerged as a consequence of the interaction of a reasonable prior with the likelihood, as implemented through Bayes rule. A similar parsimony characterizes the account of small spot colors.

The Bayesian approach provides a clear generalization path. Both models presented here have to date undergone only preliminary tests against data, and it is likely that conditions can be found that produce model failures. For example, a few classic and more numerous recent studies of color and lightness constancy have sought to generalize to stimulus conditions beyond diffusely illuminated flat matte surfaces (Hochberg & Beck, 1954; Gilchrist, 1980; Nishida & Shinya, 1998; Bloj, Kersten, & Hurlbert, 1999; Yang & Maloney, 2001; Boyaci, Maloney, & Hersh, 2003; Boyaci, Doerschner, & Maloney, 2004; Doerschner, Boyaci, & Maloney, 2004; Delahunt & Brainard, 2004a; Ripamonti et al., 2004; Motoyoshi, Nishida, Sharan, & Adelson, 2007; Xiao & Brainard, 2008; see also Fleming, Dror, & Adelson, 2003). It is not surprising that experiments of this sort reveal phenomena that cannot be accounted for by the Bayesian constancy model presented above, since that model is based on a likelihood and prior that do not allow for spatially varying illumination, objects and scenes that vary in three dimensions or for objects with geometrically complex surface reflectance functions. On the other hand, the Bayesian prescription tells us that to build more general models, it may be sufficient to generalize the likelihoods and priors. This is less daunting than developing models de novo. Recently, models that may be conceived as generalizing the likelihood function to account for directional illumination and variation in object pose have been reported (Boyaci et al., 2003; Bloj et al., 2004; Boyaci, Doerschner, & Maloney, 2006).

Although the Bayesian approach has the appealing features described above, it will not provide a complete account of vision when employed alone. First, the approach is silent about mechanism. The way in which the models are implemented on a digital computer need not speak to how equivalent performance is achieved by the nervous system. The value of a Bayesian model in understanding mechanism is indirect. The model tells us what input-output relationship any mechanistic model must satisfy.

One common misconception about Bayesian models is that because the likelihood and prior are clearly separated in the formulation, this separation should be expected in the physiological representation. That this is not necessary is easily grasped if one considers that there is a simple linear receptive field interpretation of the Bayesian model of trichromatic image reconstruction. For any choice of prior distribution, the appropriate transformation from cone responses to the estimated ideal image may be accomplished by weighted sums that link cone responses to each estimated L, M, or S value (Brainard et al., 2008). In a neural implementation of this sort, the properties of the likelihood and prior are jointly and implicitly encoded in the specific values of the weights; neither the likelihood nor the prior appears explicitly.

An often-asked question that arises in the context of Bayesian models is "Where did the priors come from?" If we view the role of the prior in a Bayesian model as specifying how the visual system resolves the ambiguity in the likelihood, then it becomes clear that this question is not really about the priors. Rather, the question is really "How did evolution and development shape the visual system so that it operates effectively in the natural environment?" This is an excellent question but one that applies to any model that correctly describes stable adult performance. The current absence of an answer is not a weakness that is specific to the Bayesian approach. Indeed, one appeal of the Bayesian approach is that the prior is represented explicitly. Thus the prior that is used in a successful model converts the raw experimental data to a form that can be compared to measurements of the statistics of the natural environment (see Brainard et al., 2006). When the derived prior matches the environmental statistics, the Bayesian model provides a quantitative link between behavioral performance and the statistics of the environment. When there is a mismatch between the derived prior and environmental statistics, the modeling emphasizes that our understanding remains incomplete.

Finally, we close by noting there are interesting and important factors that the Bayesian approach, as elaborated here, does not include. Because information processing consumes energy, for example, there may be tradeoffs between optimal estimation and efficient processing. A number of

analyses interpret mechanisms of early color vision as efficient solutions to information transmission and representation (Buchsbaum & Gottschalk, 1983; Derrico & Buchsbaum, 1990; Atick, 1992; van Hateren, 1993; Ruderman et al., 1998; Parraga, Brelstaff, Troscianko, & Moorehead, 1998; von der Twer & MacLeod, 2001; Parraga, Troscianko, & Tolhurst, 2002; Lee, Wachtler, & Sejnowski, 2002; Doi, Inui, Lee, Wachtler, & Sejnowski, 2003; Caywood, Willmore, & Tolhurst, 2004; Wachtler, Doi, Lee, & Sejnowski, 2007). Although this work shares with the Bayesian models presented here an emphasis on the statistics of the ensemble of natural scenes, it differs in its emphasis. The Bayesian models presented in this chapter focus on veridicality rather than efficiency of representation. Moreover, additional insight is likely to be obtained when both veridicality and precision of representation are considered together (see Stocker & Simoncelli, 2006; Abrams, Hillis, & Brainard, 2007; Hillis & Brainard, 2007).

ACKNOWLEDGMENTS This work was supported by NIH RO1 EY10016. I thank S. Allred, J. Nachmias, and B. Wandell for helpful comments on the manuscript.

REFERENCES

Abrams, A. B., Hillis, J. M., & Brainard, D. H. (2007). The relation between color discrimination and color constancy: When is optimal adaptation task dependent? *Neural Comput.*, *19*, 2610–2637.

Adelson, E. H. (1999). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press.

Adelson, E. H., & Pentland, A. P. (1996). The perception of shading and reflectance. In D. Knill & W. Richards (Eds.), *Visual perception: Computation and psychophysics* (pp. 409–423). New York: Cambridge University Press.

Arend, L. E., & Reeves, A. (1986). Simultaneous color constancy. *J. Opt. Soc. Am. [A]*, *3*, 1743–1751.

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing. *Network Comput. Neural Syst.s, 3*, 213–251.

Attneave, F. (1954). Some informational aspets of visual perception. *Psychol. Rev.*, *61*, 183–193.

Barrow, H. G., & Tenenbaum, J. M. (1978). Recovering intrinsic scene characteristics from images. In A R. Hanson & E. M. Riseman (Eds.), *Computer vision systems* (pp. 3–26). New York: Academic Press.

Blackwell, D., & Girschick, M. A. (1954). *Theory of games and statistical decisions*. New York: Wiley.

Bloj, A., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, *402*, 877–879.

Bloj, M., Ripamonti, C., Mitha, K., Greenwald, S., Hauck, R., & Brainard, D. H. (2004). An equivalent illuminant model for the effect of surface slant on perceived lightness. *J. Vis.*, *4*, 735–746.

Boyaci, H., Doerschner, K., & Maloney, L. T. (2004). Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity. *J. Vis.*, *4*, 664–679.

Boyaci, H., Doerschner, K., & Maloney, L. T. (2006). Cues to an equivalent lighting model. *J. Vis.*, *6*, 106–118.

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *J. Vis.*, *3*, 541–553.

Brainard, D. H. (1994). *Bayesian method for reconstructing color images from trichromatic samples*. Paper presented at the 47th Annual IS&T Meeting, Rochester, NY.

Brainard, D. H. (1998). Color constancy in the nearly natural image: 2. Achromatic loci. *J. Opt. Soc. Am. [A]*, *15*, 307–325.

Brainard, D. H. (2004). Color constancy. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (Vol. 1, pp. 948–961). Cambridge, MA: MIT Press.

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image: 1. Asymmetric matches. *J. Opt. Soc. Am. [A]*, *14*, 2091–2110.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *J. Opt. Soc. Am. [A]*, *14*, 1393–1411.

Brainard, D. H., Kraft, J. M., & Longère, P. (2003). Color constancy: Developing empirical tests of computational models. In R. Mausfeld & D. Heyer (Eds.), *Colour perception: Mind and the physical world* (pp. 307–334). Oxford, UK: Oxford University Press.

Brainard, D. H., Longere, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., & Xiao, B. (2006). Bayesian model of human color constancy. *J. Vis.*, *6*, 1267–1281.

Brainard, D. H., & Wandell, B. A. (1986). Analysis of the retinex theory of color vision. *J. Opt. Soc. Am. [A]*, *3*, 1651–1661.

Brainard, D. H., & Wandell, B. A. (1992). Asymmetric color-matching: How color appearance depends on the illuminant. *J. Opt. Soc. Am. [A]*, *9*, 1433–1448.

Brainard, D. H., Williams, D. R., & Hofer, H. (2008). Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color appearance of small spots. *J. Vis.*, *8*(5): 15, 1–23.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *J. Franklin Inst.*, *310*, 1–26.

Buchsbaum, G., & Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. R. Soc. Lond. B Biol. Sci.*, *220*, 89–113.

Burnham, R. W., Evans, R. M., & Newhall, S. M. (1957). Prediction of color appearance with different adaptation illuminations. *J. Opt. Soc. Am. [A]*, *47*, 35–42.

Burton, G. J., & Moorehead, I. R. (1987). Color and spatial structure in natural images. *Appl. Optics*, *26*, 157–170.

Caywood, M. S., Willmore, B., & Tolhurst, D. J. (2004). Independent components of color natural scenes resemble V1 neurons in their spatial and color turning. *J. Neurophysiol.*, *91*, 2859–2873.

Cohen, J. (1964). Dependency of the spectral reflectance curves of the Munsell color chips. *Psychon. Sci.*, *1*, 369–370.

Delahunt, P. B., & Brainard, D. H. (2004a). Color constancy under changes in reflected illumination. *J. Vis.*, *4*, 764–778.

Delahunt, P. B., & Brainard, D. H. (2004b). Does human color constancy incorporate the statistical regularity of natural daylight? *J. Vis.*, *4*, 57–81.

Derrico, J. B., & Buchsbaum, G. (1990). A computational model of spatiochromatic image coding in early vision. *J. Vis. Commun. Image Rep.*, *2*, 31–38.

Doerschner, K., Boyaci, H., & Maloney, L. T. (2004). Human observers compensate for secondary illumination originating in nearby chromatic surfaces. *J. Vis.*, *4*, 92–105.

Doi, E., Inui, T., Lee, T.-W., Wachtler, T., & Sejnowski, T. J. (2003). Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes. *Neural Comput.*, *15*, 397–417.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. [A]*, *4*, 2379–2394.

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *J. Vis.*, *3*, 347–368.

Geisler, W. S., & Kersten, D. (2002). Illusions, perception, and Bayes. *Nat. Neurosci.*, *5*, 508–510.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gilchrist, A. L. (2006). *Seeing black and white*. Oxford, UK: Oxford University Press.

Gilchrist, A. L. (1980). When does perceived lightness depend on perceived spatial arrangement? *Percept. Psychophys.*, *28*, 527–538.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., et al. (1999). An anchoring theory of lightness perception. *Psychol. Rev.*, *106*, 795–834.

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *J. Vis.*, *7*(4): 2, 1–15.

Helmholtz, H. (1910). *Helmholtz's physiological optics* (Translation from the 3rd German ed.). New York: Optical Society of America.

Helson, H., & Jeffers, V. B. (1940). Fundamental problems in color vision: II. Hue, lightness, and saturation of selective samples in chromatic illumination. *J. Exp. Psychol.*, *26*, 1–27.

Hillis, J. M., & Brainard, D. H. (2007). Distinct mechanisms mediate visual detection and identification. *Curr. Biol.*, *17*, 1714–1719.

Hochberg, J. E., & Beck, J. (1954). Apparent spatial arrangement and perceived brightness. *J. Exp. Psychol.*, *47*, 263–266.

Hofer, H., Singer, B., & Williams, D. R. (2005). Different sensations from cones with the same photopigment. *J. Vis.*, *5*, 444–454.

Jaaskelainen, T., Parkkinen, J., & Toyooka, S. (1990). A vector-subspace model for color representation. *J. Opt. Soc. Am. [A]*, *7*, 725–730.

Kaiser, P. K., & Boynton, R. M. (1996). *Human color vision* (2nd ed.). Washington, DC: Optical Society of America.

Katz, D. (1935). *The world of colour* (R. B. MacLeod & C. W. Fox, Trans.). London: Kegan, Paul, Trench Truber & Co.

Knill, D., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proc. Natl. Acad. Sci. USA*, *96*, 307–312.

Krantz, D. (1968). A theory of context effects based on cross-context matching. *J. Math. Psychol.*, *5*, 1–48.

Land, E. H. (1986). Recent advances in retinex theory. *Vis. Res.*, *26*, 7–21.

Lee, T.-W., Wachtler, T., & Sejnowski, T. J. (2002). Color opponency is an efficient representation of spectral properties of natural scenes. *Vis. Res.*, *42*, 2095–2103.

Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. [A]*, *3*, 1673–1683.

Maloney, L. T. (2002). Statistical decision theory and biological vision. In D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world* (pp. 145–189). New York: Wiley.

Maloney, L. T., & Wandell, B. A. (1986). Color constancy: A method for recovering surface spectral reflectances. *J. Opt. Soc. Am. [A]*, *3*, 29–33.

Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constraints. *Vis. Res.*, *41*, 2653–2668.

Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 13–36). Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

McCann, J. J., McKee, S. P., & Taylor, T. H. (1976). Quantitative studies in retinex theory: A comparison between theoretical predictions and observer responses to the "Color Mondrian" experiments. *Vis. Res.*, *16*, 445–458.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, *447*, 206–209.

Newhall, S. M., Nickerson, D., & Judd, D. B. (1943). Final report of the O.S.A. subcommittee on the spacing of Munsell colors. *J. Opt. Soc. Am. [A]*, *33*, 385–412.

Nickerson, D. (1957). *Spectrophotometric data for a collection of Munsell samples*. Washington, DC: U.S. Department of Agriculture.

Nishida, S., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *J. Opt. Soc. Am. [A]*, *15*, 2951–2965.

Parraga, C. A., Brelstaff, G., Troscianko, T., & Moorehead, I. R. (1998). Color and luminance information in natural scenes. *J. Opt. Soc. Am. [A]*, *15*, 563–569.

Parraga, C. A., Troscianko, T., & Tolhurst, D. J. (2002). Spatiochromatic properties of natural images and human vision. *Curr. Biol.*, *12*, 483–487.

Pratt, W. K. (1978). *Digital image processing*. New York: John Wiley.

Rao, R. P. N., Olshausen, B. A., & Lewicki, M. S. (Eds.). (2002). *Probabilistic models of the brain: Perception and neural function*. Cambridge, MA: MIT Press.

Ripamonti, C., Bloj, M., Mitha, K., Greenwald, S., Hauck, R., Maloney, S. I., et al. (2004). Measurements of the effect of surface slant on perceived lightness. *J. Vis.*, *4*, 747–763.

Ruderman, D. L., Cronin, T. W., & Chiao, C. C. (1998). Statistics of cone responses to natural images: Implications for visual coding. *J. Opt. Soc. Am. [A]*, *15*, 2036–2045.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularties of the terrestrial world? In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 495–532). New York: Oxford University Press.

Speigle, J. M., & Brainard, D. H. (1999). Predicting color from gray: The relationship between achromatic adjustment and asymmetric matching. *J. Opt. Soc. Am. [A]*, *16*, 2370–2376.

Stiles, W. S. (1967). Mechanism concepts in colour theory. *J. Colour Group*, *11*, 106–123.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.*, *9*, 578–585.

van Hateren, J. H. (1993). Spatiotemporal contrast sensitivity of early vision. *Vis. Res.*, *33*, 257–267.

von der Twer, T., & Macleod, D. I. A. (2001). Optimal nonlinear codes for the perception of natural colors. *Network Comput. Neural Systs.*, *12*, 395–407.

von Kries, J. (1902/1970). Chromatic adaptation. In D. L. MacAdam (Ed.), *Sources of color vision* (pp. 109–119). Cambridge, MA: MIT Press.

Wachtler, T., Doi, E., Lee, T.-W., & Sejnowski, T. J. (2007). Cone selectivity derived from the responses of the retinal cone mosaic to natural scenes. *J. Vis.*, *7*, 1–14.

Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *J. Exp. Psychol.*, *38*, 310–324.

Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat. Neurosci.*, *5*,598–604.

Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae* (2nd ed.). New York: Wiley.

Xiao, B., & Brainard, D. H. (2008). Surface gloss and color perception of 3D objects. *Vis. Neurosci.*, *25*, 371–385.

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vis. Res.*, *41*, 2581–2600.

28 Wiring of Receptive Fields and Functional Maps in Primary Visual Cortex

DARIO L. RINGACH

ABSTRACT This chapter deals with the question of how receptive fields and cortical maps are wired before the onset of the classical critical period. In particular, we ask how simple-cell receptive fields may arise without an intermediate phase of overlap between ON and OFF subregions and without the need for correlated spontaneous activity in the developing thalamus. We discuss one possible solution, the statistical connectivity hypothesis, which postulates that initial wiring of the cortex is highly constrained by the spatial arrangement of the retinal ganglion cell mosaic and their coverage ratios. We examine a recently confirmed prediction of the theory: that orientation bandwidth must depend on the location of neurons within the orientation map. Finally, the theory is shown to predict the existence of orientation scotomas: At any given retinal location, not all orientations can be represented equally well by neurons in primary visual cortex.

Wiring of receptive fields and functional maps in primary visual cortex

To fully understand the development and the adult organization of primary visual cortex, there are three separate questions that must be addressed. First, we need to discover the mechanisms that are responsible for wiring receptive fields and cortical maps at the earliest stages of development (Albus & Wolf, 1984; Hubel & Wiesel, 1963; Sherk & Stryker, 1976). Second, we need a description of how activity-dependent processes maintain, modify, or refine these initial structures during the critical period (Crair, Gillespie, & Stryker, 1998; Crowley & Katz, 2002; Katz & Crowley, 2002; Miller, Erwin, & Kayser, 1999; Swindale, 1996). Third, we need to understand which features of the resulting receptive fields and cortical maps are vital for normal visual processing and which may arise as an epiphenomenon of developmental processes and wiring constraints (Adams & Horton, 2003; Chklovskii & Koulakov, 2000; Horton & Adams, 2005; Koulakov & Chklovskii, 2001; Purves, Riddle, & Lamantia, 1992; Swindale, 1991). This chapter deals primarily with the early establishment of the circuit. We want to know how receptive fields and cortical maps are wired before the onset of the critical period. The reason this question is a critical piece of the puzzle of V1 development is clarified by a summary of some experimental findings.

In their pioneering studies of visual cortex, Hubel and Wiesel (1963) demonstrated that kittens lacking normal visual experience have cells that are tuned for orientation. They also found that orientation-tuned cells cluster into populations of similar preference, suggesting the presence of an early orientation map in these young animals (Hubel & Wiesel, 1963). These early cortical responses were also found to be heavily dominated by contralateral input (Crair et al., 1998; Fregnac & Imbert, 1978; Movshon & Van Sluyters, 1981).

Recent studies using intrinsic imaging of cortical activity in combination with single-unit electrophysiology have confirmed and refined these classical findings, showing that in kittens, orientation maps and ocular dominance columns are already present by two weeks of age (Crair et al., 1998; Crair, Horton, Antonini, & Stryker, 2001). Furthermore, there are no obvious differences in the development of ocular dominance and orientation maps of normal and binocularly deprived animals up to the third postnatal week, demonstrating that normal visual stimulation is not necessary for the early wiring of receptive fields and maps (Crair et al., 1998, 2001).

An important finding is that receptive fields with segregated ON and OFF subregions, which are characteristic of simple cells, are observed in the thalamo-recipient layers 4 and 6 as soon as the cortex becomes visually responsive (Albus & Wolf, 1984; Blakemore & Van Sluyters, 1975; Braastad & Heggelund, 1985; Hubel & Wiesel, 1963; Sherk & Stryker, 1976). Furthermore, the ratio between the numbers of simple and complex cells in the first weeks of development remains approximately constant and does not differ from that in the adult (Albus & Wolf, 1984; Braastad & Heggelund, 1985).

DARIO L. RINGACH Department of Neurobiology and Psychology, Jules Stein Eye Institute, David Geffen School of Medicine, University of California, Los Angeles, California

These data are difficult to reconcile with the notion that simple-cell receptive fields develop from a set of heavily overlapping ON/OFF inputs (Linsker, 1986; Miller, 1994; Reid & Alonso, 1995). If this were the case, one would have predicted (1) an initial prevalence of receptive fields with overlapping ON/OFF receptive fields with a progressive spatial segregation of subregions during development and (2) an increase in the ratio of simple to complex cells during this developmental process. Instead, the available data indicate that salient features of the adult cortical organization, including the subregion segregation of simple cells, orientation, and ocular dominance maps, manifest themselves at the earliest stages of cortical development, well before the onset of the critical period.

We are thus faced with the challenge of explaining how receptive fields and cortical maps are wired initially. A couple of hypotheses have been considered so far. One possibility is that the presence of structured spontaneous activity in the developing thalamus could drive the initial thalamocortical wiring (Miller, 1994; Miller et al., 1999). Such correlation-based models predict a specific pattern of activity: Thalamic cell pairs having the same center sign (either ON/ON or OFF/OFF) should be more correlated than cells with different center signs at small distances (on the scale of a subregion width); an opposite pattern, in which same-sign cells are less correlated than opposite sign pairs, should be observed at larger separations. This pattern of spontaneous activity and a synaptic connectivity rule by which "neurons that fire together, wire together" ensure the emergence of segregated ON/OFF subregions (simple-cell receptive fields) from overlapping ON/OFF inputs. To guarantee the periodicity of orientation columns, an additional mechanism that leads nearby cells to develop similar receptive fields and cells at large separations to develop different receptive fields must be invoked.

Recent measurements of spontaneous activity in the developing thalamus, however, have failed to corroborate the predicted pattern of correlations (Ohshiro & Weliky, 2006; Weliky & Katz, 1999). Instead of the predicted Mexican hat profile, one observes a Gaussian falloff of correlation for same-sign receptive fields and zero correlation for different-sign pairs at all distances. Under these conditions, the model fails to develop segregated ON/OFF subregions. These ideas could still be rescued by invoking a more complex "split constraint" that conserves the synaptic strength of ON and OFF center cells separately during development. However, its biological implementation is hard to imagine (Ohshiro & Weliky, 2006).

A second possibility is that molecular cues, involved in axonal guidance/patterning, help to establish the initial cortical architecture (Crowley & Katz, 2000; Katz & Crowley, 2002). If thalamic afferents carrying signals from overlapping ON/OFF-center receptive fields are to be sorted, perhaps this is done by different molecular markers specifying the locations where each input type is allowed to create synaptic contacts on the target neuron, thereby generating nonoverlapping ON/OFF subregions.

Molecular guidance has been shown to be involved in the establishment of a coarse retinotopy and in retinogeniculate laminar segregation (Cang, Kaneko, Yamada, Woods, & Stryker, 2005; Huberman, 2007), yet its role in guiding connectivity at the fine spatial scales required to shape the structure of the subregions in individual receptive fields has never been demonstrated and appears unlikely. For example, it is difficult to conceive how different simple cells, on the same orientation column, could coordinate the expression of markers so that all their receptive fields develop similar orientation preferences.

Molecular patterning has also been proposed as underlying the generation of ocular dominance columns (Crowley & Katz, 1999; Hubener & Bonhoeffer, 1999; Katz & Crowley, 2002). This may be a more appealing possibility, owing to the larger spatial scales involved, but we should consider that functional maps are related in specific ways. In the cat, for example, orientation pinwheels tend to align with the centers of ocular dominance domains (Bartfeld & Grinvald, 1992; Grinvald, Frostig, Siegel, & Bartfeld, 1991), and peaks of low/high spatial frequency domains tend to align with the pinwheel centers (Everson, 1998, Issa, Trepel, & Stryker, 2000). Envisioning how molecular guidance by itself could simultaneously explain the development of cortical maps (retinotopy, ocular dominance, orientation, spatial frequency) and their relationships appears to be a rather difficult task indeed.

Arguably, these considerations weaken the case for spontaneous activity in the developing thalamus and molecular guidance as explanations for the early establishment of the cortical architecture. While it is premature to rule out their involvement altogether, one cannot help but wonder whether there are any other wiring mechanisms that have not been considered.

The proposal that I would like to discuss here was born out of the realization that the common assumption that simple cells develop from a set of overlapping ON/OFF receptive fields is not supported by the available data. Thus asking how simple cells arise from overlapping inputs is *not* the right question to pursue. The relevant question is how the subregions of simple cells could be wired without going through a developmental phase of substantial ON/OFF overlap.

The answer to this question, I propose, is that receptive fields of LGN afferents are not expected to have a high degree of overlap in the first place (Ringach, 2004, 2007). This assertion is based on the known statistics of retinal ganglion cell mosaics, the degree of overlap of their receptive fields (coverage ratios), and the fact that LGN cells are dominated by a single ganglion cell input (thereby reflecting the

spatial statistics of the RGC mosaic and their receptive fields) (Chichilnisky & Kalmar, 2002; Cleland & Lee, 1985; Usrey, Reppas, & Reid, 1999; Wassle, Boycott, & Illing, 1981). In other words, the solution to the problem of simple-cell wiring might rest on the recognition that the problem does not actually exist.

At first, this might appear to be a trivial observation, but it turns out to be one with profound implications. In a formal development of these ideas, I have shown that they not only lead to an explanation for how simple-cell receptive fields emerge early during development, but also provide an account for statistics of monosynaptic connections between the thalamus and cortex, the relationships between various cortical maps, and the dependence of neuronal selectivity across the functional architecture of the cortex. In the remainder of this chapter, we review some of these results at the conceptual level and highlight some curious predictions of the model.

Statistical connectivity theory

The working hypothesis behind the proposal is that the *blueprint* for the formation of simple-cell receptive fields in V1, the feature maps in the cortex and several of their mutual relationships, resides in the layout of the retinal ganglion cell mosaics in the contralateral eye combined with a simple statistical connectivity scheme between the thalamus and the cortex.

The notion that the structure of the RGC mosaic could influence the development of receptive fields in the cortex was first formulated by Wassle, Boycott, & Illing, (1981, p. 192) and considered by Soodak (1987). These investigators noted that nearest neighbors on the X-cell RGC mosaic tend to be of opposite sign, thus generating ON/OFF pairs in close proximity. The spatial statistics of the RGC mosaic are reflected in the LGN, as thalamic cells are driven mostly by one input. If cortical cells were to pool in space a small number of nearby thalamic afferents representing neighboring RGC inputs, the result would be the sum of slightly displaced ON and OFF Gaussian receptive fields. This would generate a simple-cell receptive field with a preference for orientation (figure 28.1*A*).

It should now be apparent that this idea provides a way out of the puzzle of simple-cell wiring early in development, as it explains both the reason for a lack of overlapping ON/OFF inputs and how a simple (isotropic) pooling of afferent inputs can generate orientation-tuned cells. The model also predicts that individual subregion widths should be comparable to the width of the receptive field center of LGN afferents, as has been found experimentally (Alonso, Usrey, & Reid, 2001; Hubel & Wiesel, 1959, 1962). Furthermore, the model is consistent with the notion that simple-cell receptive fields in layer 4 are initially formed by summation of inputs from a relatively small number of spatially displaced ON- and OFF-center LGN afferents (Gilbert, 1977; Heggelund, 1986).

If, as postulated, the local distribution of RGC receptive fields is responsible for inducing orientation tuning in the cortex, then it must be the case that V1 cells with different orientation preferences must get their input from RGC with receptive fields that do not overlap significantly (otherwise, the resulting preferred orientations would be similar) (figure 28.1*B*). This predicts the presence of *orientations scotomas*: At any given retinal location, and when input is restricted to one type of ganglion cells, not all orientations can be represented equally well by V1 receptive fields. This is because the RGC mosaic imposes severe constraints on the receptive fields that can be constructed at each location.

Experimental evidence for the existence of orientation scotomas would be of high significance, as it would not only confirm the model's prediction but also invalidate the classic view of the cortical hypercolumn as a population of neurons with overlapping receptive fields tuned for possible orientations preferences. The finding that large orientation differences between nearby cells are accompanied by large jumps in the retinotopic location is consistent with the prediction (Das & Gilbert, 1997), but we must note that this result remains highly controversial (Bosking, Crowley, & Fitzpatrick, 2002; Buzas, Volgushev, Eysel, & Kisvarday, 2003). Future studies using in vivo two-photon imaging (Ohki, Chung, Ch'ng, Kara, & Reid, 2005; Ohki et al., 2006) along with precise receptive field mapping procedures could potentially settle this issue.

The existence of orientation scotomas could potentially be tested psychophysically. One such attempt was performed by Zanker and Braitenberg (1996), who measured performance in an orientation discrimination task using very small lines of different lengths at various eccentricities. They noted that when detection performance is expressed in terms of the cortical size of the line (by taking into account the average human magnification factor), all data points collapse into a single curve, as is typical of many psychophysical tasks (Rovamo & Virsu, 1979; Watson, 1987).

Notably, it was found that performance levels of 60–80% can be achieved with stimuli that cover merely 0.2–0.3 mm of cortical territory. Because these are smaller than the size of an entire orientation cycle (about 1 mm in primates), they decided to use such small stimuli to map biases in orientation discrimination performance at different retinal locations. This attempt was only partially successful; in individual sessions, they obtained orientation biases that appeared to resemble orientation maps as measured by optical means. Unfortunately, these maps could not be consistently reproduced from one experimental session to the next. An obvious difficulty with their experiment (which the authors acknowledged) was the lack of control over eye movements and the

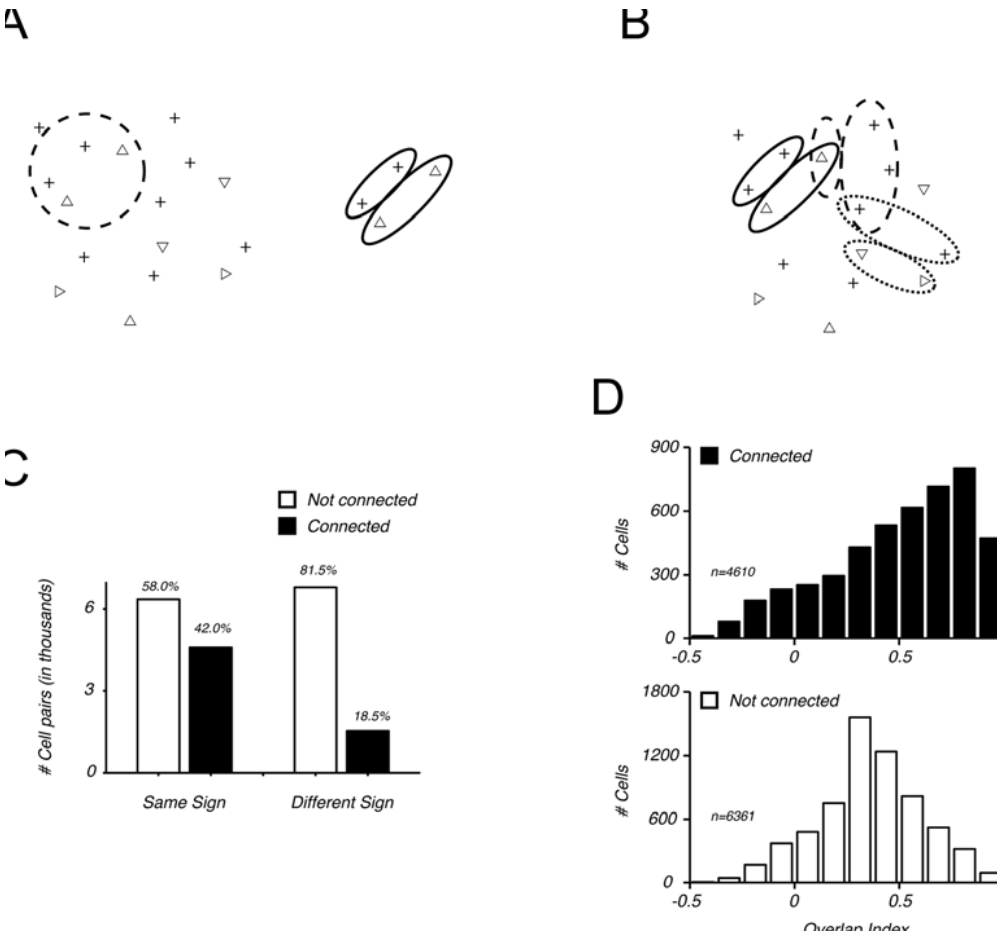


FIGURE 28.1 Conceptual description of statistical connectivity and some of its consequences. (*A*) The theory posits that the distribution of ON-center (plus signs) and OFF-center (triangles) retinal ganglion cells receptive fields, along with a moderate coverage ratio (the solid disks indicate 1 standard deviation of the receptive field center), and the isotropic sampling of incoming afferents (dashed circle) are responsible for the establishment of the early cortical architecture. In this example, sampling from the afferents within the indicated area would generate a receptive field with adjacent ON/OFF subregions are shown to the right. (*B*) Two consequences of statistical connectivity can be inferred from this simple diagram. First, to obtain receptive fields with substantially different orientation, one must move to a different retinal location (orientation scotomas). Second, there should be a tendency for overlapping simple-cell receptive fields to have the same sign within the overlap area. (*C*, *D*) The theory is consistent with the statistics of thalamocortical connectivity. Both the sign rule and the distribution of receptive field overlap is explained by the model.

ability to repeatedly stimulate the exact same retinal location across experimental sessions. We are currently attempting to perform similar psychophysical experiments while carefully monitoring eye movements.

Another success of statistical connectivity is in explaining the probability and strength of monosynaptic connections from thalamus to cortex. In particular, the model replicates the *sign rule* of connectivity (Alonso et al., 2001; Reid & Alonso, 1995). This refers to the finding that the probability of a monosynaptic connection is highest when the geniculate receptive field overlaps a simple-cell subregion of the same signature (either ON or OFF), while the probability of "inappropriate" connections between receptive fields of opposite signature is much lower. These data have been interpreted as supporting the existence of precise rules of synaptic connectivity in accordance with the classic wiring scheme for simple cells (Hubel & Wiesel, 1962). However, this interpretation rests on the assumption that the cortex receives afferents from a large number of overlapping ON and OFF geniculate receptive fields, which, as was discussed above, is incorrect.

Statistical connectivity offers an explanation for the apparent precision of thalamocortical wiring (figure 28.1*C*). In the model, an ON subregion of a simple-cell results in the event that an ON-center rather than OFF-center geniculate cell dominates that location of visual space. As a consequence, one would expect a tendency for ON subregions to avoid OFF inputs. The reason is simply that there are no OFF inputs to avoid at that location. A similar analysis is to plot the distribution of the correlation coefficient between the spatial receptive field of thalamic afferents and overlapping cortical receptive fields for cases in which cells were

connected or not. The model replicates two salient properties of these data: (1) Connected receptive fields have higher overlap indices than nonconnected pairs, and (2) the overlap index in the nonconnected case has a positive mean. The latter one is of high significance, as it indicates that one cannot experimentally find pairs with high negative correlation (whether connected or not). The only possible explanation is that such pairs are nonexistent.

Statistical connectivity also predicts orientation columns. That is, even if cells in the same column sample *independently* from the LGN afferents with some probability of connection, there is a tendency for them to have the same orientation preference (Ringach, 2004). This statistical sampling generates a diversity of the *relative* phase of the receptive fields, spanning the entire range from even symmetric to odd symmetric reception fields (DeAngelis, Ghose, Ohzawa, & Freeman, 1999). However, the model predicts that overlapping subregions of simple-cell receptive fields must agree in sign (figure 28.1*B*). This observation leads to another testable prediction of the model: different ON or OFF afferents from the LGN dominate the cortical input at different locations, generating a spatial-phase map on the cortex. Recent data confirmed this prediction in cat cortex (Jin et al., 2008), with similar organizations described previously in ferrets (Zahs & Stryker, 1988) and minks (McConnell & Levay, 1984).

Statistical connectivity and cortical maps

Statistical connectivity generates not only simple-cell receptive fields and orientations columns, but orientation maps as well (Ringach, 2007; Soodak, 1987). Detailed computer simulations have revealed several other important consequences of statistical connectivity (Ringach, 2007). First, the theory predicts that the local magnification factor should be directly related to local fluctuations in the density of the RGC mosaic. That is, fluctuations in the density of RGCs around its mean value, on a small spatial scale, should induce a parallel fluctuation in the local cortical magnification factor. The existence of fluctuations in the local magnification factor could potentially be tested by using in vivo two-photon imaging along with careful receptive field mapping. Second, statistical connectivity predicts that the orientation tuning width of individual cells should depend on their location within the orientation map. As we discuss below, we have recently obtained new data confirming this relationship. Third, it predicts the existence of spatial frequency maps and a tendency for preference for extreme spatial frequencies (high or low) to align with orientation pinwheels, as reported experimentally (Issa et al., 2000). Fourth, it predicts the existence of clustered regions of broad selectivity (identified as putative cytochrome-oxidase blobs) and a tendency for pinwheels to lie preferentially in interblob regions, which is supported by extant data (Bartfeld & Grinvald, 1992; Hubener, Shoham, Grinvald, & Bonhoeffer, 1997).

The relationship between orientation selectivity and the location of neurons within the map has been the subject of several studies. While the subthreshold tuning of cells conforms with the prediction of the statistical connectivity model with broader tuning seen near pinwheel locations (Marino et al., 2005), the tuning of *spike* responses has been described as being invariant across the map (Maldonado, Godecke, Gray, & Bonhoeffer, 1997; Marino et al., 2005). We decided to take another look at this question using a novel technique (Nauhaus & Ringach, 2007). The method involves measuring the orientation map with optical means and then recording from multiple neurons with a micromachined electrode array. Briefly, we obtained the orientation map on a cortical patch of V1 using optical imaging (Bonhoeffer & Grinvald, 1991; Grinvald & Hildesheim, 2004) (figure 28.2*A*). Subsequent to the acquisition of the orientation maps, we implanted a 10×10 electrode array with grid spacing of 400 mm on the same cortical patch. Finally, we measured the orientation tuning of neuronal responses across the array using reverse correlation in the orientation domain, where the stimulus consists of a rapid sequence of gratings at random orientation (Ringach, Hawken, & Shapley, 1997). The average orientation-triggered response, at the optimal time delay, generates an orientation tuning curve (figure 28.2*B*). A Gaussian fit to each tuning curve provides an estimate of the preferred orientation, θ_0, and tuning width, $\Delta\theta$.

The position of the array with respect to the orientation map is estimated by searching for the location that yields the maximum agreement between the preferred orientations measured optically and those measured from the electrode array. The outcome of this computation in one experiment is shown in the scatterplot of figure 28.2*C*, along with the estimated location of the array in figure 28.2*A*. The location of the array on the surface of the cortex is determined by just three parameters: two for its translation and one for its rotation. Because the number of parameters is small in comparison with the number of electrodes, the optimization problem is greatly overdetermined and robust to noise.

Once the array location has been estimated, we computed a measure of homogeneity in orientation preference for the neighborhood surrounding each electrode. We define a local homogeneity index at a cortical point by computing the magnitude of a vector sum. The vector angles are determined by the orientations in the map, and the magnitudes are defined by a spatial two-dimensioned Gaussian window centered at the given cortical point. The local homogeneity index is bounded between 0 and 1. It is high in regions that exhibit similar preferences for orientation, such as iso-orientation domains, and low in regions where the preferences for orientation are diverse, such as near pinwheel

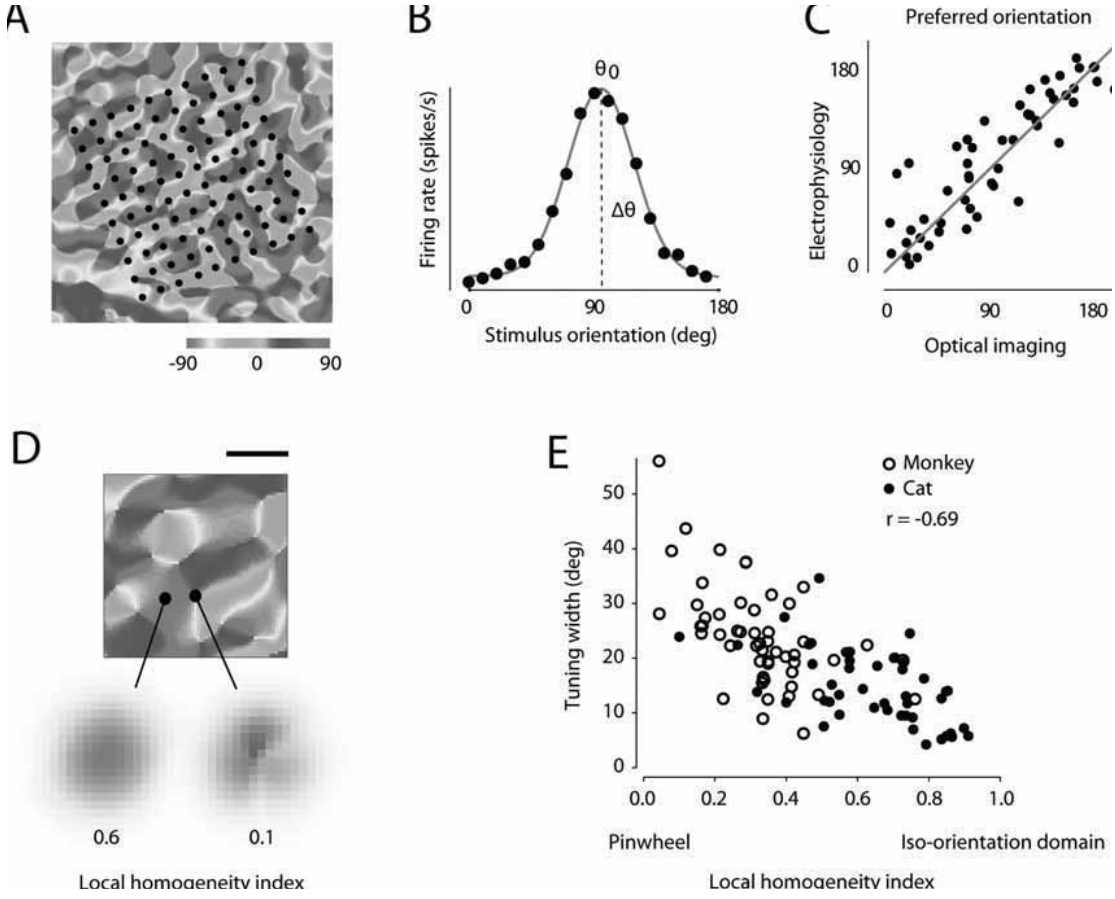


FIGURE 28.2 Orientation tuning bandwidth and local map structure. (*A*) Example of an orientation preference map in macaque visual cortex along with the recovered location of the microelectrode array. (*B*) Reverse correlation in the orientation domain (Ringach et al., 1997) was used to measure the tuning curves at each electrode site simultaneously. The example here shows the average spike rate triggered to the presentation of each orientation in a rapid stimulus sequence, yielding a preferred orientation, θ_0, and tuning width, $\Delta\theta$. (*C*) The estimated location of the array (solid dots in panel *A*) was estimated by finding the optimal translation/rotation parameters for which the preferred orientations as measured via reverse correlation matched those measured optically. The scatterplot illustrates the optimal correlation in one instance. (*D*) A local homogeneity index was defined to capture the diversity of orientation preferences around each cortical point. The example illustrates two locations with a low homogeneity index of 0.1 attained near a pinwheel and a location with a high index of 0.6 in an iso-orientation domain. (*E*) Spatial distribution of the local homogeneity index for the same patch of cortex as the one shown in panel *D*. (*F*) Isolation of single units. Only units that could be very well isolated, as is typical of the principal component analysis here, were used in our analyses of tuning bandwidth and local map structure. (See color plate 42.)

centers (figure 28.2*E*). Using this method, we found that orientation tuning width and homogeneity index are negatively correlated in both monkeys ($r = -0.56$, $p = 0.00001$) and cats ($r = -0.56$, $p = 0.00005$) (figure 28.2*E*), as predicted by the model.

It should be emphasized that statistical connectivity is not the only possible explanation for this trend. One likely contribution to this relationship comes from the fact that the tuning properties of the local environment of a cell is likely to determine the tuning of the intracortical feedback signal and, in turn, the tuning of the cell (Marino et al., 2005; McLaughlin, Shapley, & Shelley, 2003; Schummers, Marino, & Sur, 2002). These two explanations are not mutually exclusive.

Testing statistical connectivity

The status of statistical connectivity as a viable working hypothesis for the early wiring of receptive field and cortical maps derives from the fact that it is an extremely simple concept that can explain a large set of data, including the structure and emergence of simple receptive fields, the relationship between cortical maps, and the dependence of neuronal selectivity across functional maps.

There are many predictions that remain to be tested. A particularly interesting one is the existence of orientation scotomas. However, there are other ways to test the theory directly. If the theory is correct, given the structure of the RGC mosaic in the contralateral eye, one should be able

to predict the structure of the orientation map in the cortex. A positive result would unequivocally show a relationship between these two structures.

One way to proceed is to image cortical maps while recording in the LGN from a location with receptive fields that provide input to the imaged cortical region (Reid & Alonso, 1995). Injection of a retrograde label and subsequent recovery could allow the reconstruction of the RGC mosaic at the same location (Wassle, Boycott, & Illing, 1981). Careful measurement of the magnification factor is necessary to relate the RGC structure to the spatial scale of the orientation map. Ideally, such an experiment should be performed in animals in which the ipsilateral eye has been enucleated and the animal has been reared with only monocular input, as input from the ipsilateral eye can change the original structure of the orientation maps (Crair et al., 1998; Farley, Yu, Jin, & Sur, 2007).

Concluding remarks

In this chapter, I have tried to emphasize some important puzzles that remain unanswered about the early development of the visual system. I explained how in an effort to address these issues, in particular the rapid emergence of simple cells, we were led to a simple idea about how constraints imposed by the distribution of RGCs could establish the early blueprint of receptive fields and maps in the cortex. It was gratifying to discover that these ideas had already been considered in prior studies (Soodak, 1987; Wassle, Peichl, & Boycott, 1981). My contribution to the topic has been to extend the theory to include probabilistic sampling of the RGC mosaic and better understand the implications of these ideas beyond the establishment of orientation maps in the cortex, including the statistics of thalamocortical connectivity, the relationship between different cortical maps, and the dependence of tuning across the functional maps of the cortex. It is worth noting that similar concepts have been invoked as potential explanations for other properties of visual cells, such as their chromatic tuning (Lennie & Movshon, 2005; Solomon & Lennie, 2007).

There are a couple of obvious directions in which the theory can develop. First, if these ideas can be confirmed experimentally for the case of monocular input, it would be important to develop extensions to binocular inputs and investigate what mechanisms are involved in the wiring of binocular simple-cell receptive fields. In this respect, I have speculated that fluctuations in the RGC densities of matching retinal locations as possible seeding the pattern of ocular dominance columns (Ringach, 2007). Second, it is now clear that some rodents develop orientation tuning and simple cells without the presence of an orientation map (Ohki et al., 2005; Van Hooser, Heimel, Chung, Nelson, & Toth, 2005). Investigating the applicability and limitations of statistical connectivity to other species that lack functional maps can be expected to be an instructive exercise.

As a final observation, I would like to submit the notion that statistical connectivity offers a potential explanation for the evolutionary emergence of simple-cell receptive fields. Given that edgelike filters are a crucial element in the representation of natural images (Bell & Sejnowski, 1997; Olshausen & Field, 1996; Simoncelli & Olshausen, 2001), one may ask what evolutionary path could have led to the appearance of simple-cell receptive fields in so many diverse species. Statistical connectivity suggests a simple recipe with only two ingredients: ON/OFF center cells with appropriate coverage ratios and simple spatial summation of the output of such neurons. Then the emergence of simple-cell receptive fields appears almost unavoidable, and their explicit signaling of object boundaries could have endowed organisms with a survival advantage. Such a route to the development of simple-cell receptive fields seems sensible, and it does not require the evolution of precise connectivity rules or the prior development of complex patterns of spontaneous activity and the learning rules that are required to implement them.

REFERENCES

Adams, D. L., & Horton, J. C. (2003). Capricious expression of cortical columns in the primate brain. *Nat. Neurosci.*, *6*, 113–114.

Albus, K., & Wolf, W. (1984). Early post-natal development of neuronal function in the kitten's visual cortex: A laminar analysis. *J. Physiol. Lond.*, *348*, 153–185.

Alonso, J. M., Usrey, W. M., & Reid, R. C. (2001). Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *J. Neurosci.*, *21*, 4002–4015.

Bartfeld, E., & Grinvald, A. (1992). Relationships between orientation-preference pinwheels, cytochrome-oxidase blobs, and ocular-dominance columns in primate striate cortex. *Proc. Natl. Acad. Sci. USA*, *89*, 11905–11909.

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vis. Res.*, *37*, 3327–3338.

Blakemore, C., & Van Sluyters, R. C. (1975). Innate and environmental factors in development of kitten's visual cortex. *J. Physiol. Lond.*, *248*, 663–716.

Bonhoeffer, T., & Grinvald, A. (1991). Iso-orientation domains in cat visual-cortex are arranged in pinwheel-like patterns. *Nature*, *353*, 429–431.

Bosking, W. H., Crowley, J. C., & Fitzpatrick, D. (2002). Spatial coding of position and orientation in primary visual cortex. *Nat. Neurosci.*, *5*, 874–882.

Braastad, B. O., & Heggelund, P. (1985). Development of spatial receptive-field organization and orientation selectivity in kitten striate cortex. *J. Neurophysiol.*, *53*, 1158–1178.

Buzas, P., Volgushev, M., Eysel, U. T., & Kisvarday, Z. F. (2003). Independence of visuotopic representation and orientation map in the visual cortex of the cat. *Eur. J. Neurosci.*, *18*, 957–968.

Cang, J., Kaneko, M., Yamada, J., Woods, G., Stryker, M. P., & Feldheim, D. A. (2005). Ephrin-as guide the formation of functional maps in the visual cortex. *Neuron*, *48*, 577–589.

Chichilnisky, E. J., & Kalmar, R. S. (2002). Functional asymmetries in ON and OFF ganglion cells of primate retina. *J. Neurosci.*, *22*, 2737–2747.

Chklovskii, D. B., & Koulakov, A. A. (2000). A wire length minimization approach to ocular dominance patterns in mammalian visual cortex. *Physica A Statist. Mechanics Appl.*, *284*, 318–334.

Cleland B. G., & Lee, B. B. (1985). A comparison of visual responses of cat lateral geniculate-nucleus neurons with those of ganglion-cells afferent to them. *J. Physiol. Lond.*, *369*, 249–268.

Crair, M. C., Gillespie, D. C., & Stryker, M. P. (1998). The role of visual experience in the development of columns in cat visual cortex. *Science*, *279*, 566–570.

Crair, M. C., Horton, J. C., Antonini, A., & Stryker, M. P. (2001). Emergence of ocular dominance columns in cat visual cortex by 2 weeks of age. *J. Comp. Neurol.*, *430*, 235–249.

Crowley, J. C., & Katz, L. C. (1999). Development of ocular dominance columns in the absence of retinal input. *Nat. Neurosci.*, *2*, 1125–1130.

Crowley, J. C., & Katz, L. C. (2000). Early development of ocular dominance columns. *Science*, *290*, 1321–1324.

Crowley, J. C., & Katz, L. C. (2002). Ocular dominance development revisited. *Curr. Opin. Neurobiol.*, *12*, 104–109.

Das, A., & Gilbert, C. D. (1997). Distortions of visuotopic map match orientation singularities in primary visual cortex. *Nature*, *387*, 594–598.

DeAngelis, G. C., Ghose, G. M., Ohzawa, I., & Freeman, R. D. (1999). Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons. *J. Neurosci.*, *19*, 4046–4064.

Farley, B. J., Yu, H., Jin, D. Z., & Sur, M. (2007). Alteration of visual input results in a coordinated reorganization of multiple visual cortex maps. *J. Neurosci.*, *27*, 10299–10310.

Fregnac, Y., & Imbert, M. (1978). Early development of visual cortical-cells in normal and dark-reared kittens: Relationship between orientation selectivity and ocular dominance. *J. Physiol. Lond.*, *278*, 27–44.

Gilbert, C. D. (1977). Laminar differences in receptive-field properties of cells in cat primary visual-cortex. *J. Physiol. Lond.*, *268*, 391–421.

Grinvald, A., Frostig, R. D., Siegel, R. M., & Bartfeld, E. (1991). High-resolution optical imaging of functional brain architecture in the awake monkey. *Proc. Natl. Acad. Sci. USA*, *88*, 11559–11563.

Grinvald, A., & Hildesheim, R. (2007). VSDI: A new era in functional imaging of cortical dynamics. *Nat. Rev. Neurosci.*, *5*, 874–885.

Heggelund, P. (1986). Quantitative studies of the discharge fields of single cells in cat striate cortex. *J. Physiol. Lond.*, *373*, 277–292.

Horton, J. C., & Adams, D. L. (2005). The cortical column: A structure without a function. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *360*, 837–862.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol. Lond.*, *148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in cat's visual cortex. *J. Physiol. Lond.*, *160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *J. Neurophysiol.*, *26*, 994–1002.

Hubener, M., & Bonhoeffer, T. (1999). Eyes wide shut. *Nat. Neurosci.*, *2*, 1043–1045.

Hubener, M., Shoham, D., Grinvald, A., & Bonhoeffer, T. (1997). Spatial relationships among three columnar systems in cat area 17. *J. Neurosci.*, *17*, 9270–9284.

Huberman, A. D. (2007). Mechanisms of eye-specific visual circuit development. *Curr. Opin. Neurobiol.*, *17*, 73–80.

Issa, N. P., Trepel, C., & Stryker, M. P. (2000). Spatial frequency maps in cat visual cortex. *J. Neurosci.*, *20*, 8504–8514.

Jin, J. Z., Weng, C., Yeh, C. I., Gordon, J. A., Ruthazer, E. S., Stryker, M. P., et al. (2008). On and off domains of geniculate afferents in cat primary visual cortex. *Nat. Neurosci.*, *11*, 88–94.

Katz, L. C., & Crowley, J. C. (2002). Development of cortical circuits: Lessons from ocular dominance columns. *Nat. Rev. Neuroscience 3*, 34–42.

Koulakov, A. A., & Chklovskii, D. B. (2001). Orientation preference patterns in mammalian visual cortex: A wire length minimization approach. *Neuron 29*, 519–527.

Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). *J. Opt. Soc. Am. [A]*, *22*, 2013–2033.

Linsker, R. (1986). From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proc. Natl. Acad. Sci. USA*, *83*, 7508–7512.

Maldonado, P. E., Godecke, I., Gray, C. M., & Bonhoeffer, T. (1997). Orientation selectivity in pinwheel centers in cat striate cortex. *Science*, *276*, 1551–1555.

Marino, J., Schummers, J., Lyon, D. C., Schwabe, L., Beck, O., Wiesing, P., et al. (2005). Invariant computations in local cortical networks with balanced excitation and inhibition. *Nat. Neurosci.*, *8*, 194–201.

McConnell, S. K., & Levay, S. (1984). Segregation of on-center and off-center afferents in mink visual-cortex. *Proc. Natl. Acad. Sci. USA*, *81*, 1590–1593.

McLaughlin, D., Shapley, R., & Shelley, M. (2003). Large-scale modeling of the primary visual cortex: Influence of cortical architecture upon neuronal response. *J. Physiol. Paris*, *97*, 237–252.

Miller, K. D. (1994). A model for the development of simple cell receptive-fields and the ordered arrangement of orientation columns through activity-dependent competition between on- and off-center inputs. *J. Neurosci.*, *14*, 409–441.

Miller, K. D., Erwin, E., & Kayser, A. (1999). Is the development of orientation selectivity instructed by activity? *J. Neurobiol.*, *41*, 44–57.

Movshon, J. A., & Van Sluyters, R. C. (1981). Visual neural development. *Annu. Rev. Psychol.*, *32*, 477–522.

Nauhaus, I., & Ringach, D. L. (2007). Precise alignment of micromachined electrode arrays with V1 functional maps. *J. Neurophysiol.*, *97*, 3781–3789.

Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P., & Reid, R. C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, *433*, 597–603.

Ohki, K., Chung, S. Y., Kara, P., Hubener, M., Bonhoeffer, T., & Reid, R. C. (2006). Highly ordered arrangement of single neurons in orientation pinwheels. *Nature*, *442*, 925–928.

Ohshiro, T., & Weliky, M. (2006). Simple fall-off pattern of correlated neural activity in the developing lateral geniculate nucleus. *Nat. Neurosci.*, *9*, 1541–1548.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Purves, D., Riddle, D. R., & Lamantia, A. S. (1992). Iterated patterns of brain circuitry (or how the cortex gets its spots). *Trends Neurosci.*, *15*, 362–368.

Reid, R. C., & Alonso, J. M. (1995). Specificity of monosynaptic connections from thalamus to visual-cortex. *Nature*, *378*, 281–284.

Ringach, D. L. (2004). Haphazard wiring of simple receptive fields and orientation columns in visual cortex. *J. Neurophysiol.*, *92*, 468–476.

Ringach, D. L. (2007). On the origin of the functional architecture of the cortex. *PLoS ONE*, 2, e251.

Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, *387*, 281–284.

Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Exp. Brain Res.*, *37*, 495–510.

Schummers, J., Marino, J., & Sur, M. (2002). Synaptic integration by V1 neurons depends on location within the orientation map. *Neuron*, *36*, 969–978.

Sherk, H., & Stryker, M. P. (1976). Quantitative study of cortical orientation selectivity in visually inexperienced kitten. *J. Neurophysiol.*, *39*, 63–70.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.*, *24*, 1193–1216.

Solomon, S. G., & Lennie, P. (2007). The machinery of colour vision. *Nat. Rev. Neurosci.*, *8*, 276–286.

Soodak, R. E. (1987). The retinal ganglion-cell mosaic defines orientation columns in striate cortex. *Proc. Natl. Acad. Sci. USA*, *84*, 3936–3940.

Swindale, N. V. (1991). Coverage and the design of striate cortex. *Biol. Cybern.*, *65*, 415–424.

Swindale, N. V. (1996). The development of topography in the visual cortex: A review of models. *Network Comput. Neural Syst.*, *7*, 161–247.

Usrey, W. M., Reppas, J. B., & Reid, R. C. (1999). Specificity and strength of retinogeniculate connections. *J. Neurophysiol.*, *82*, 3527–3540.

Van Hooser, S. D., Heimel, J. A. F., Chung, S., Nelson, S. B., & Toth, L. J. (2005). Orientation selectivity without orientation maps in visual cortex of a highly visual mammal. *J. Neurosci.*, *25*, 19–28.

Wassle, H., Boycott, B. B., & Illing, R. B. (1981). Morphology and mosaic of on-beta and off-beta cells in the cat retina and some functional considerations. *Proc. R. Soc. Lond. B Biol. Sci.*, *212*, 177–195.

Wassle, H., Peichl, L., & Boycott, B. B. (1981). Morphology and topography of on-alpha and off-alpha cells in the cat retina. *Proc. R. Soc. Lond. B Biol. Sci.*, *212*, 157–175.

Watson, A. B. (1987). Estimation of local spatial scale. *J. Opt. Soc. Am. [A]*, *4*, 1579–1582.

Weliky, M., & Katz, L. C. (1999). Correlational structure of spontaneous neuronal activity in the developing lateral geniculate nucleus in vivo. *Science*, *285*, 599–604.

Zahs, K. R., & Stryker, M. P. (1988). Segregation of ON and OFF afferents to ferret visual cortex. *J. Neurophysiol.*, *59*, 1410–1429.

Zanker, J. M., & Braitenberg, V. (1996). Psychophysical mapping of orientation sensitivity in the human cortex. In A. Aertsen & V. Braitenberg (Eds.), *Brain theory: Biological basis and computational principles* (pp. 19–36). Amsterdam: Elsevier Science.

29 Encoding and Decoding with Neural Populations in the Primate Cortex

EYAL SEIDEMANN, YUZHI CHEN, AND WILSON S. GEISLER

ABSTRACT Environmental stimuli are encoded by large neural populations in sensory cortical areas and subsequently decoded into motor plans by large neural populations in motor cortical areas. Furthermore, large populations of neurons are likely to exhibit new emergent properties that are difficult or impossible to infer from the activity of single neurons, recorded one at a time. Thus, to understand encoding and decoding in the cortex, it is essential to measure and analyze neural population responses, ideally in behaving subjects. In this chapter we review recent progress in experimental techniques for measuring simultaneously the activity of large neural populations; we discuss several important emergent properties of neural population responses; and we describe a Bayesian ideal observer framework that, when applied to simultaneous measurements of neural population responses and behavioral performance, can be used to rigorously explore encoding and decoding strategies at the level of neural populations.

Sensory stimuli are encoded by large populations of neurons in sensory cortical areas and later decoded into motor plans that are implemented by large populations of neurons in motor cortical areas (figure 29.1). The specific encoding and decoding circuits that are implemented in the cortex are undoubtedly shaped by a number of factors, including the natural tasks that the organism performs, the properties of the sensory stimuli and musculature that are relevant to performing those tasks, the biophysical and anatomical properties of neurons, and the available space and metabolic resources.

These considerations raise many fundamental questions. For example, what are the number and identity of neurons in a given cortical area that contribute to a behavioral response? What aspects of the stimulus or behavior are represented in the signals from these neurons, and how are they encoded in the neural response? How are these signals combined over space and time to mediate behavior? Are these encoding and decoding algorithms fixed, or do they depend on the specific requirements of the task? Addressing such questions is one of the key challenges facing systems neuroscience.

Over the past several decades, the primary approach to addressing these questions has been single-neuron electrophysiology in combination with measurement of the stimuli and/or behavioral responses. While much has been learned by using this approach, it will be difficult, if not impossible, to fully understand how neural circuits in the mammalian cortex encode and decode information on the basis of single-unit electrophysiology. The reason is simply that large populations of neurons are likely to exhibit new and fundamentally different *emergent properties* that might not be evident from recordings of individual neurons, one at a time. Therefore we believe that it is important to shift the focus from single neurons to populations of neurons by directly measuring the properties of neural population responses, ideally in behaving subjects. In addition, to understand the implications of these properties, it is necessary to develop a theoretical framework that would allow a rigorous exploration of possible encoding and decoding strategies at the level of neural populations. In this chapter, we describe recent progress along these two lines of research.

The main focus in this review is on the primary visual cortex (V1) of primates, because this sensory area is arguably the best understood in terms of its anatomy, neuronal response properties, and functional organization (de Valois & de Valois, 1988; Hubel & Weisel, 1977). However, most of the results and theoretical considerations described below are likely to apply to other cortical areas and species. To illustrate some of the general experimental and theoretical issues that are relevant for understanding population coding and decoding, we discuss some of our measurements with voltage-sensitive dye imaging (VSDI) in V1, because these measurements forced us to think generally about the expected response properties of large populations of neurons. Two goals of this chapter are to stimulate more research on population encoding and decoding and to emphasize the importance of using multiple complementary techniques to measure neural population activity.

EYAL SEIDEMANN, YUZHI CHEN, and WILSON S. GEISLER Department of Psychology and Center for Perceptual Systems, University of Texas, Austin, Texas

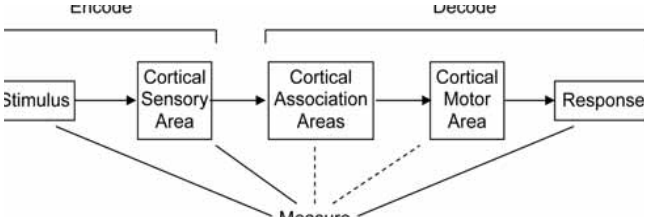


FIGURE 29.1 Schematic representation of processing stages in perceptual tasks. The solid lines indicate the measurements that are the focus of this review.

We begin this review by describing a computational framework that can be used to explore possible population encoding and decoding mechanisms. We then describe the advantages and disadvantages of some of the experimental tools currently available for monitoring population responses in vivo. We next discuss key properties of population responses in the primate cortex. We end with a general discussion that includes open questions and future research directions.

Theoretical framework for studying population encoding and decoding

The responses of the neural population in a given sensory cortical area contain a certain amount of information that is potentially available to support performance in a given task. This information could be characterized, in principle, by measuring the statistical relationship between relevant environmental stimuli and response properties of the neural population at the given stage. Further, this information can be quantified, in principle, by deriving the ideal Bayesian observer (for the given task) that has complete knowledge of the statistical relationship between environmental stimuli and the population response. An ideal observer is a theoretical device that performs a task optimally given the available input signals, knowledge of the prior probabilities of different possible stimuli, and knowledge of the cost and benefits of the possible stimulus-response outcomes (Geisler, 1989; Green & Swets, 1966). The performance of the ideal Bayesian observer is the appropriate measure of the neural information potentially available for specific tasks, and it is the measure we will use here.[1]

Not all response properties of a neural population might be relevant for a given task (i.e., would improve the ideal observer's performance); therefore one important goal is to determine which properties of the population's responses carry information relevant to a given task and which do not. On the other hand, there may be relevant response properties that the organism does not use to perform a given task because of limitations of the decoding mechanisms. For example, subsequent stages might be unable to select responses from an individual neuron (or arbitrary subset of neurons) in the population or use precise spike-timing information, even if this could lead to improved performance. It is also possible that the organism uses some irrelevant (potentially performance-degrading) properties. Thus a second important goal is to determine which properties of the population responses the organism uses to perform a given task. Finally, a third important goal is to determine how these properties of the population responses are translated by subsequent circuits into behavioral responses. Here we define the *actual code* for a task as those specific properties of the population responses that the organism does use in performing the task, we define the *actual encoder* for the task as those specific neural mechanisms that translate the sensory stimulus into the actual code, and we define the *actual decoder* for the task as the specific subsequent neural mechanisms that translates the actual code into perceptual decisions and motor plans (see figure 29.1).

For any given task and set of neural constraints (e.g., some fixed number of neurons with specified anatomical and biophysical limits), there is an *optimal encoder* that translates the sensory stimuli into the neural population code that carries the most information relevant to performing the task. Similarly, for any given task and population code, there is an *optimal decoder*. The concepts and mathematics of ideal observer theory can be used to derive both optimal encoders and decoders. Determining the optimal encoder and decoder can be very useful in the quest to identify the actual encoder and decoder, because the exercise generally leads to a deep understanding of the computational requirements of the task and often provides principled (and sometimes unexpected) hypotheses for the neural mechanisms.

Although much has been learned and much remains to be learned about encoding mechanisms in sensory areas (e.g., Geisler, 2008; Simoncelli & Olshausen, 2001), because of space limitations the focus in this chapter is on evaluating possible decoding mechanisms based on measured neural population responses in sensory areas.

To further illustrate these ideas, consider a thought experiment in which an organism is required to discriminate between two barely discriminable stimuli. Assume that all the sensory information relevant for performance in this task passes through a single cortical area and that, as experimenters, we have precise access to the responses of all neurons in this area, to the stimulus, and to the behavioral response of the subject. In this case, the Bayesian framework outlined above allows us to determine how to perform the task optimally based on these neural signals and to determine the behavioral sensitivity that could be supported by this optimal decoder. Because we have access to all the neural signals that are available for the organism to perform the task, the optimal decoder must do as well as, or better than, the organism.

The ideal Bayesian observer analysis provides several key benefits. Equal performance of the ideal observer and the organism implies that the organism is using all the relevant

neural information and is employing an optimal decoder; it also implies that subsequent processing stages are effectively noiseless. Significantly better performance for the ideal observer implies that the organism is using either a suboptimal code, a suboptimal decoder or both. (Note that a decoder could be suboptimal because it implements the wrong algorithm and/or because it introduces neural noise.) If the optimal decoder is significantly more sensitive than the organism, one can explore the range of possible codes and decoders that could still account for the subject's performance, assuming noiseless subsequent processing stages. To determine what aspects of the neural responses are essential in order to explain the performance of the organism, one can measure the consequences of ignoring some aspects of the neural signals on the performance of the ideal observer. For example, one can evaluate the importance of the precise timing of spikes by integrating over a temporal window that can be gradually increased to determine when the ideal observer's performance falls short of the subject's. Similarly, one can evaluate the importance of individual neurons by only considering the summed response from pools of m neighboring neurons for various values of m. Any suboptimal code that reduces the sensitivity of the optimal decoder to a level that is lower than the sensitivity of the organism can be rejected. More generally, any combination of code and decoder that falls short of the subject's performance can be rejected as a possible mechanism used by the organism.

While the above thought experiment clearly cannot be achieved in the primate cortex by any single currently available technique (see discussion), we demonstrate below that this framework can still be useful when applied in conjunction with existing neurophysiological methods.

Current techniques for measuring neural population responses in vivo

Currently available techniques for monitoring responses of neural populations fall far short of fulfilling our thought experiment. Here, we briefly review the advantages and limitations of three techniques that seem particularly promising for studying population coding in the primate cortex: multielectrode recording, optical imaging with voltage sensitive dyes, and two-photon imaging of calcium-indicator signals in multiple single cortical neurons.

Large arrays of microelectrodes for extracellular recordings from multiple single cortical neurons are an important emerging technique (e.g., Churchland, Yu, Sahani, & Shenoy, 2007; Kelly et al., 2007; Nauhaus, Benucci, Carandini, & Ringach, 2008; Nicolelis & Ribeiro, 2002). The advantages of this technique are submillisecond temporal resolution and access to spiking activity at the single-neuron level. The main disadvantage is that it samples the activity fairly coarsely over space (typically a few hundred microns between electrodes, due to electrode density limitations). Therefore it is limited to recording a tiny fraction of the neurons in several square millimeters. This technique may also suffer from sampling biases (e.g., bias toward neurons with large cell bodies and therefore large spikes, bias against very selective neurons with very low spontaneous rate). The spatial sampling of this technique could potentially be improved by also recording multiunit activity and local field potentials from each electrode, but the quantitative relationship between these measurements and the activity of single units still needs to be explored.

VSDI has recently been adapted for studies in alert, behaving primates (Seidemann, Arieli, Grinvald, & Slovin, 2002; Slovin, Arieli, Hildesheim, & Grinvald, 2002). In this technique, voltage-sensitive fluorescence dyes are applied topically to the surface of the cortex (for a review, see Grinvald & Hildesheim, 2004). The dyes bind nonselectively to cellular membranes. In the membrane, the dye molecules transduce, essentially instantaneously and linearly, changes in membrane potential into fluorescence signals. The dye signals therefore represent the summed changes in membrane potentials in cell bodies, axons, and dendrites in the superficial cortical layers. The main advantages of this technique are its millisecond temporal resolution and its large field of view (typically 1–2 cm^2). As will be discussed below, this large field of view is necessary in order to cover the entire region within a given cortical area that is activated even by a small localized stimulus (figure 29.2). The main limitation of this technique is its spatial resolution. Because dendrites and axons of cortical neurons spread horizontally over several hundred microns, signals collected by each imaging pixel combine responses from large numbers of neurons and neural elements in a nonselective manner. This nonselective summation could significantly degrade the VSDI signals relative to the actual neural code. A second limitation is that the signals are restricted to the superficial cortical layers (approximately the top 500 μm). However, the superficial layers contain many apical dendrites of neurons whose cell bodies are located in deeper layers. Finally, VSDI signals in vivo are weak and contaminated by various sources of noise. However, our recent results discussed below, suggest that most of the nonneural sources of noise can be removed reliably (Chen, Geisler, & Seidemann, 2006, 2008).

Two-photon imaging of calcium-indicator signals is an exciting new addition to the arsenal of tools that are available for studying population responses in vivo (Kerr et al., 2007; Ohki, Chung, Ch'ng, Kara, & Reid, 2005; Stosiek, Garaschuk, Holthoff, & Konnerth, 2003; for a recent review, see Kerr & Denk, 2008). This is the only technique that allows measuring the activity of all the neurons in a small region simultaneously, with single-neuron resolution. Because changes in the calcium signal are relatively slow, inferring activity at the level of single spikes

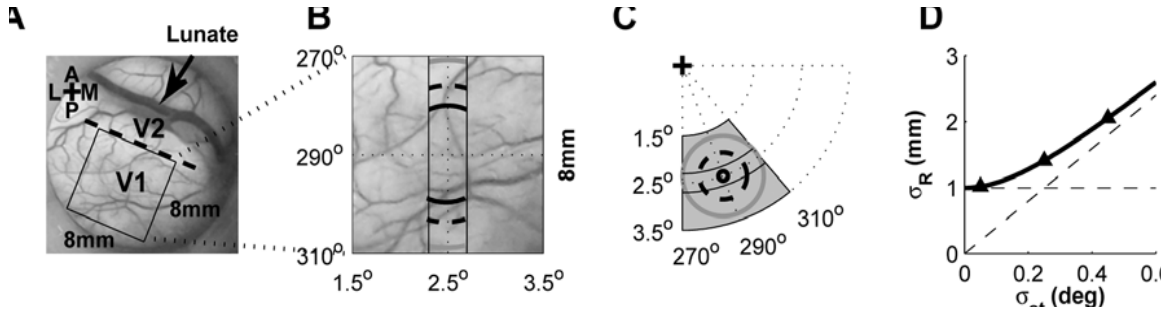


FIGURE 29.2 Expected spread of activity in the visual cortex in response to a small localized visual stimulus. (*A*) Cranial window over V1 in the left hemisphere of one monkey. The cortical vasculature is seen through a transparent artificial dura (Arieli, Grinvald, & Slovin, 2002). A typical region of interest of 8×8 mm^2 with its anterior border running along the V1/V2 border is indicated by the black square. (*B*) Expanded view of the cortical vasculature in the 8×8 mm^2 region of interest. (*C*) Representation of the lower right visual field with a fixation crosshair in the top left. The shaded wedge region is the approximate portion of the visual field that is represented in the patch in panel *B*. The mapping from visual space to the cortex is indicated in panels *C* and *B*, respectively. The solid, dashed, and gray circles in panel *C* represent the outline (at $\pm 2\sigma_{st}$) of Gabor patches with σ_{st} of 0.05°, 0.25°, and 0.45°, respectively. The corresponding arcs in panel *B* indicate the expected spread of activity in response to the three stimuli in a narrow strip of cortex along the representation of 2.5° eccentricity (see text for additional details). (*D*) Expected spread of cortical activity σ_R as a function of stimulus size σ_{st}. The dashed horizontal line indicates the minimal spread, which corresponds to the average V1 receptive field size (σ_{rf} of 0.25 degree) multiplied by the CMF at this eccentricity (4 mm/degree). The oblique dashed line shows the expected spread of cortical activity based solely on the CMF.

can be challenging, particularly when firing rates are above a few hertz. Also, because this is a scanning technique, there is an inherent tradeoff between frame rate and field of view. At high frame rates, the technique is currently limited to recording several dozen neurons within a fraction of a square millimeter. Finally, as with VSDI, this technique is currently limited to recording from the superficial cortical layers. Overall, this is a promising new technique, but more work is necessary before it will be applicable to alert, behaving primates.

These techniques promise new and exciting discoveries in the coming years. Given their limitations, however, we believe that to address questions of encoding and decoding by populations of neurons, it will be necessary to use complementary techniques and to develop quantitative understanding of the relationship between measurements provided by these and other techniques.

Properties of population responses in the primate cortex—mean response

Here, we focus on three key properties of the mean population response—spatial spread, sparseness, and temporal dynamics—using the primary visual cortex (V1) as an example.

SPATIAL SPREAD OF THE POPULATION RESPONSE Consider the spread of activity in a small patch of 8×8 mm^2 cortex on the dorsal portion of macaque V1 (figure 29.2). V1 contains a topographic map of visual space, with disproportional representation of the center of gaze (fovea). As one moves from the representation of the fovea toward the periphery (increasing the eccentricity), receptive fields (RFs) of V1 neurons become larger, and the cortical magnification factor (CMF), the distance in cortex that corresponds to a given distance in visual space, decreases, both changing approximately by a power law (Tootell, Switkes, Silverman, & Hamilton, 1988; Van Essen, Newsome, & Maunsell, 1984; Yang, Heeger, & Seidemann, 2007). This patch of cortex represents a wedged-shaped region in visual space (shaded region in figure 29.2*C*), extending approximately from an eccentricity of 1.5 degrees to 3.5 degrees (degrees of visual angle) and representing directions about the visual axis between 270 and 310 degrees (angular degrees). Here, we consider the population response in a narrow vertical strip that is centered on the cortical representation of 2.5 degrees eccentricity, where the CMF is approximately 4 mm/degree. RFs of V1 neurons have an envelope that is approximately a two-dimensional Gaussian, with a space constant σ_{rf} that is on average around 0.25 degree at this eccentricity (Nienborg, Bridge, Parker, & Cumming, 2004; Palmer, Cheng, & Seidemann, 2007). The dashed circle in figure 29.2*C* shows a typical RF with diameter of 1 degree ($4 \times \sigma_{rf}$).

The expected spatial profile of the population response in V1 can be obtained by filtering (convolving) the retinotopic projection of the stimulus to the cortex with the average RF expressed in millimeters of cortex (under linearity assumptions, which are approximately true for small localized stimuli). Specifically, if the stimulus has a Gaussian envelope (e.g., Gabor patch), the expected spread of activity in the cortex would also be a Gaussian with a space constant σ_R given by

$$\sigma_R = \sqrt{(\sigma_{st} \cdot CMF)^2 + (\sigma_{tf} \cdot CMF)^2}$$

where σ_{st} is the space constant of the Gabor patch. Thus the spread of the cortical response depends on the size of the stimulus and on the size of the average RF.

Figure 29.2*B* shows the expected extent of cortical activity in response to the stimuli with Gaussian envelopes indicated by the circles in figure 29.2*C* (plotted at diameter of $4 \times \sigma_{st}$). Finally, figure 29.2*D* shows σ_R as a function of σ_{st}. For small stimuli, σ_R asymptotes at $\sigma_{tf} \times CMF = 1$ mm. This simple analysis therefore shows that the smallest activated area in this region of V1 has a diameter of about 4 mm (see McIlwain, 1986, for a related discussion).

In the analysis above, we made several simplifying assumptions. In reality, the spread is likely to be higher, because at any given location in V1, the RF centers are significantly scattered (DeAngelis, Ghose, Ohzawa, & Freeman, 1999). In addition, because there is a wide range of RF sizes at each V1 location, with neurons with larger RFs selective to lower spatial frequencies (De Valois, Albrecht, & Thorell, 1982), the extent of the cortical spread is likely to depend on the spatial frequency of the stimulus. Nevertheless, even for the stimulus that produces the smallest spread, the response is still likely to encompass multiple square millimeters in V1.

So far, we have considered the expected spread of spiking activity in V1 based on retinotopy and average receptive field size. Figure 29.3*A* shows the spread of VSDI signals in a similar patch of macaque V1 in response to a small Gabor stimulus (σ_{st} = 0.33 degrees). As expected, activity spreads over a wide region and is well fitted by a two-dimensional Gaussian with space constants that are on the order of 1.5 to 2 mm. The response is elongated, owing to the anisotropy in the retinotopical map in this region of V1 (Tootell et al., 1988; Van Essen et al., 1984; Yang et al., 2007). Although response amplitude depends strongly on target contrast, the spatial profile of the response does not (Chen, Geisler, & Seidemann, 2006). The white ellipse shows the $\pm 2\sigma_R$ contour based on the fitted VSDI response. The black ellipse shows the $\pm 2\sigma_R$ contour that is expected on the basis of the stimulus size and the approximate CMF and average RF size in this area. The predicted σ_R is approximately 30% smaller than the observed σ_R. This difference is expected because of the subthreshold contribution to the VSD signal and because of scatter in receptive centers and sizes. The relatively small difference between the predicted spread of spiking activity and the observed VSDI spread suggests a close relationship between VSDI signals and spiking population responses.

The theoretical and experimental results that are discussed in this section indicate that even the most localized sensory stimulus activates millions of neurons in a region encompassing multiple square millimeters in sensory cortex. Next we consider how neurons in this large population are likely to respond to small localized stimuli.

Sparseness of Population Responses V1 neurons are tuned to multiple stimulus dimensions, including position, size, orientation, spatial frequency, temporal frequency, stereoscopic depth, wavelength, and motion direction (de Valois & de Valois, 1988; DeAngelis et al., 1999; Geisler & Albrecht, 1997). Intuitively, because V1 neurons are tuned across multiple dimensions, any small localized stimulus will activate strongly only a small fraction of the neurons in the population.

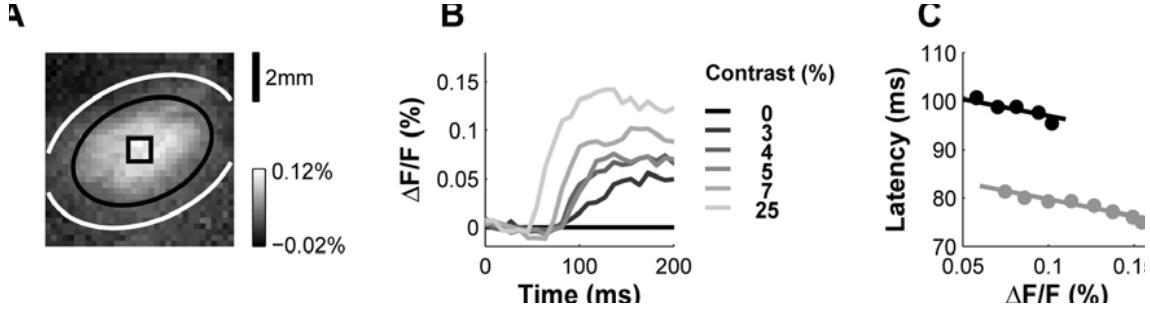


Figure 29.3 Spatiotemporal properties of V1 population response to a small Gabor target (σ_{st} of 0.33 degree) as measured by VSDI in one experiment. (*A*) Spatial profile of response amplitude to the Gabor target at 25% contrast. Response amplitude is computed as the average amplitude in a 200-ms-long temporal interval following target onset. The white ellipse indicates the contour of a two-dimensional Gaussian fit to the evoked response (at 2-standard-deviations). The black ellipse indicates the 2-standard-deviation contour of the expected region of spiking activity (see text). The square shows a 1×1 mm^2 region centered at the most sensitive location. (*B*) Time courses of average VSDI signals in response to targets at different contrasts after subtraction of the average response in target absent trials. (*C*) Response latency as a function of response amplitude for 25% (gray) and 7% (black) target contrasts. Time courses were averaged in regions with similar response amplitude and fitted with sigmoidal function. To obtain regions with similar response amplitude, the fitted two-dimensional Gaussian was divided into 10 elliptical annuli containing response amplitudes within 10 quantiles (e.g., the second innermost annulus contains location with response between 80% and 90% of R_{max}). Latency is time to half maximum. Lines show best fit linear regression.

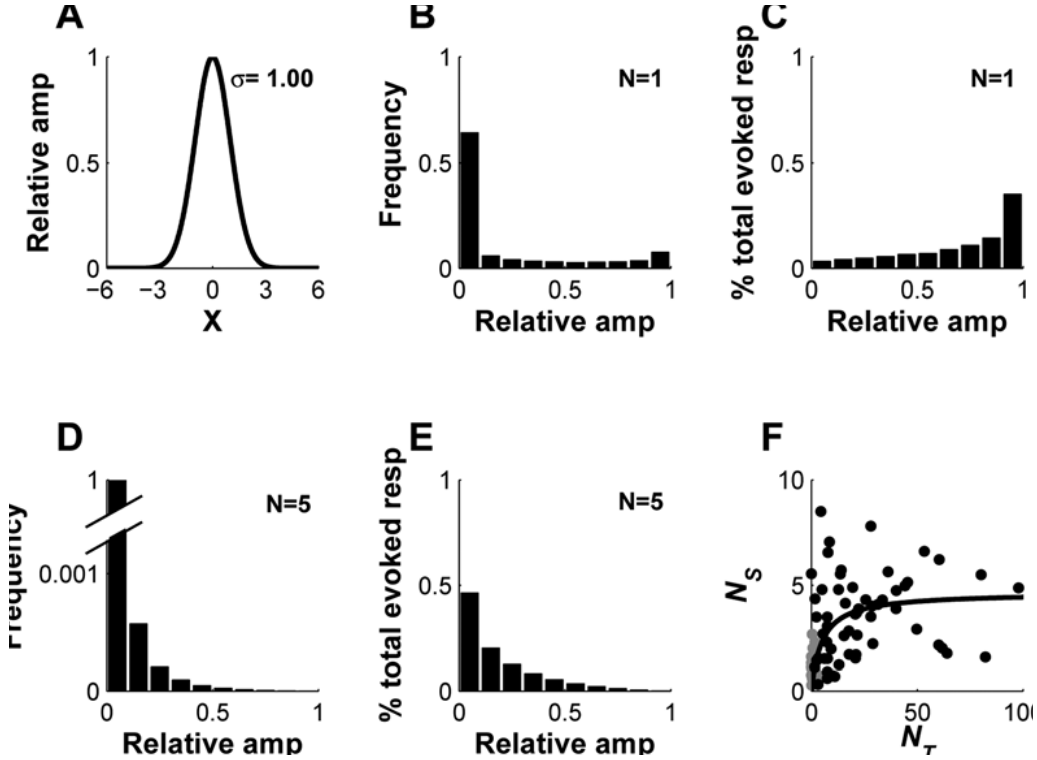


FIGURE 29.4 Interactions between tuning width, number of stimulus dimensions to which neurons are selective, and sparseness of the population response, based on a hypothetical population of neurons. (*A*) Gaussian tuning curve of one neuron across one stimulus dimension. Neurons in the population are assumed to have the same tuning curve properties across all stimulus dimensions and to uniformly cover the full stimulus range. (*B*) Frequency histogram of the relative response amplitude to a random stimulus assuming that neurons are tuned to one stimulus dimension. (*C*) Percentage of the total stimulus-evoked response contributed by neurons in the different bins in panel *B*. (*D* and *E*) Same as panels *B* and *C* but for five independent stimulus dimensions. (*F*) Quantitative relationship between baseline and stimulus-evoked response of single neurons and multiunits measured in macaque V1. The scatterplot shows the equivalent number of selective single units, N_S, that can account for the multiunit response evoked by an optimal Gabor patch versus the expected total number of single units, N_T, that contribute to the baseline multiunit response. Gray circles represent single units; black circles represent multiple units. The solid curve is the fit to the observed multiunit data with a saturating function. (See Palmer, Cheng, & Seidemann, 2007, for additional details.)

This section considers the relationship between neural tuning width, the number of stimulus dimensions that are represented by the population, and the sparseness of the response within the population. For simplicity, we ignore the specific details of the tuning properties of V1 neurons and consider a hypothetical population of neurons tuned uniformly across n stimulus dimensions. Figure 29.4*A* shows a Gaussian tuning curve of one neuron across one circular stimulus dimension (such as orientation) with 2σ equal to one-sixth of the full range, a value that is comparable to the average orientation tuning width in V1 (Geisler & Albrecht, 1997). Assume for the moment that this is the only stimulus dimension along which the neurons in the population are tuned and that the neurons uniformly cover the full stimulus range. Given the tuning curve, we can determine the fraction of neurons in the population that are expected to respond at any level of activity to a random stimulus. Figure 29.4*B* shows a frequency histogram of the expected proportion of the population at each of 10 response-level quantiles. The distribution is bimodal, with more than 50% of the neurons responding at less than 10% of their maximal response (R_{max}) but a significant fraction of neurons responding at more than 90% of R_{max}. From this distribution, we can also determine the percentage of the total stimulus-evoked population response contributed by neurons at each of the 10 response quantiles (figure 29.4*C*). In this case, about 30% of the total evoked response is contributed by the most active neurons, and the percentage of the total response decreases monotonically with decreasing mean response. Finally, we can compute ω_1, the ratio of the average stimulus-evoked response to R_{max}. In this case, ω_1 is about 0.2. (More generally, this ratio is given by $\omega_n = \omega_1^n$, where n is the number of independent stimulus dimensions.)

The picture changes dramatically if we consider five stimulus dimensions (figures 29.4*D* and 29.4*E*). Now more than 99.9% of the neurons in the population fall in the lowest amplitude quantile, and the proportion of neurons in the highest quantile is less than 10^{-6}. Similarly, when we con-

sider the percentage of the total stimulus-evoked response across the population, almost half of the response is contributed by the neurons in the lowest quantile, and less than 8% of the response is contributed by neurons responding above 50% of their R_{max}. Finally, in this case, ω_5 is equal to $\omega_1^5 \cong 3 \times 10^{-4}$. In other words, a population of 10,000 neurons is expected to produce a stimulus-evoked response that is equivalent to only three neurons firing at their R_{max}.

These theoretical considerations are qualitatively consistent with findings from a recent study in which we compared the responses of single and multiple units in V1 to small Gabor patches (figure 29.4*F*; see also Palmer et al., 2007). When the baseline of the multiunit activity is consistent with as many as 100 single neurons, the stimulus-evoked response is consistent, on average, with only four to five neurons firing at their maximal rate, indicating that most of the neurons that are contributing to the multiunit signal are only weakly responding to the stimulus.

Dynamics of the Population Response The large field of view and high temporal resolution of VSDI allow precise measurement of the spatiotemporal dynamics of the population response. Figure 29.3*B* shows the time course of the VSD response, at the center of the active region, to a Gabor patch at different contrasts. As the contrast is lowered, response amplitude drops and response latency increases. Similar interactions between stimulus contrast and response latency have been observed in single-unit electrophysiology (e.g., Albrecht, Geisler, Frazor, & Crane, 2002).

The spatiotemporal dynamics of the average stimulus-evoked response are illustrated in figure 29.3*C*. For a given contrast, response latency is almost constant across space. For example, the latency of the response to the 25% contrast Gabor patch at the peak location was only 5.7 ms faster than the latency several millimeters away at a region with response amplitudes that are only 20–30% of the peak response. Although response amplitude and latency strongly depend on target contrast, this rapid spatial spread appears to be similar at the two contrasts (for additional details, see Chen, Geisler, & Seidemann, 2008).

Properties of population responses in the primate cortex: Response variability

We now describe several key properties of the response variability of neural populations as measured by VSDI in V1 and discuss possible reasons why these properties are markedly different from those measured at the single neuron level.

Additive Versus Multiplicative Noise In single cortical neurons, the variance of the spike count during a short interval is proportional to the mean (Geisler & Albrecht, 1997; Tolhurst, Movshon, & Dean, 1983). What is the expected relationship between the mean and the variance of the response in large populations of neurons?

As was noted in the previous section, the vast majority of the neurons in the population are expected to produce a very weak response to any given stimulus. V1 neurons, however, have a significant baseline or spontaneous response even in the absence of any sensory stimulus. Because all the neurons in the population contribute to the baseline response, the variability in the population response is likely to be dominated by the variability of the baseline response, which is, by definition, stimulus independent (for additional details, see Chen et al., 2008). These considerations have two important implications. First, they imply that additive noise is expected in any technique that measures the responses of large populations of cortical neurons. Second, they imply that any decoding mechanism that pools indiscriminately the activity of a large number of neurons is likely to encounter a largely stimulus-independent additive noise.

Figure 29.5*A* shows the mean and standard deviation of the VSDI signals as a function of stimulus contrast. Consistent with the theoretical considerations discussed above, the variability of the population response is contrast independent and additive (see also Arieli, Sterkin, Grinvald, & Aertsen, 1996; Chen et al., 2006).

Two additional factors specific to VSDI may contribute to the additive nature of the observed variability. First, VSDI signals are likely to contain a significant contribution from subthreshold synaptic activity. Intracellular measurements in anesthetized cats suggest that the variability at the level of the membrane voltage may be relatively stimulus independent or even decrease with increasing target contrast (Priebe & Ferster, 2008). Second, it is possible that nonneural sources of noise dominate VSDI measurements. This is unlikely because VSDI signals in anesthetized cat are highly correlated with intracellular measurements from single neurons and with local field potentials recorded simultaneously, and are also correlated with spiking activity in the same area (Arieli et al., 1996; Grinvald et al., 1999). In addition, our results (discussed below) demonstrate that the VSDI measurements are exceedingly sensitive in simple visual detection tasks, suggesting that nonneural sources of noise are unlikely to be large.

In summary, simple theoretical considerations predict that large populations of neurons should exhibit variability that is approximately additive and not multiplicative, consistent with VSDI results. Additional measurements using complementary techniques should be used to verify this important prediction.

Spatial Correlations in the Population Response The magnitude and extent of spatial correlations in response variability can have a large impact on the improvement in performance that can be attained by pooling responses over

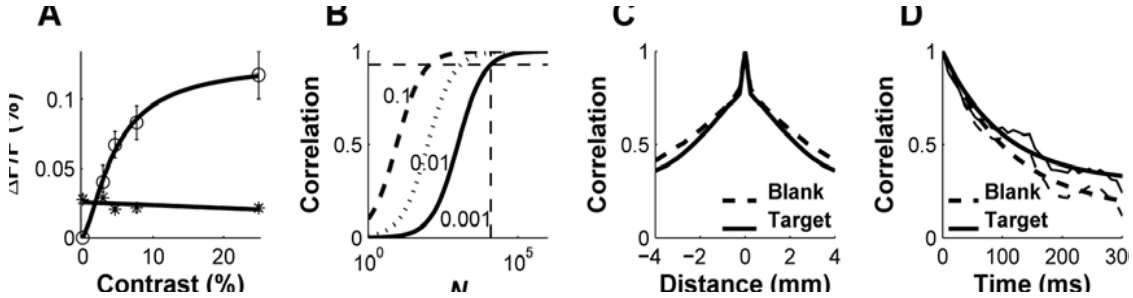


FIGURE 29.5 Statistical properties of population response variability as measured by VSDI. *A*, mean (circles) and standard deviation (asterisks) of response amplitude as a function of stimulus contrast averaged across eight VSDI experiments. Error bars indicate the standard error of the mean. The mean response as a function of contrast is fitted with a Naka-Rashton function; the standard deviation as a function of contrast is fitted with linear regression. The slope of the regression is not significantly different from zero (i.e., stimulus-independent additive noise). *B*, Expected correlations between the summed activity in two pools of neurons with uniform pairwise correlations, as a function of the number of neurons in each pool (see Chen, Geisler, & Seidemann, 2006). The value of the pairwise correlation is indicated near each curve. The dashed vertical line is the approximate number of neurons contributing to each location. The dashed horizontal line is the predicted correlation in VSDI for two locations that are 0.25 mm apart. *C*, Average correlation between two locations in one VSDI experiment as a function of the separation between the locations. *D*, Average temporal correlations between responses in two frames as a function of their separation in time. Smooth curves are exponential fits.

large populations of neurons (e.g., Abbott & Dayan, 1999; Averbeck, Latham, & Pouget, 2006; Johnson, 1980; Snippe & Koenderink, 1992; Sompolinsky, Yoon, Kang, & Shamir, 2001). Extracellular recording studies have measured the correlations in spiking activity between pairs of nearby cortical neurons (e.g., Bair, Zohary, & Newsome, 2001; Gawne & Richmond, 1993; Lee, Port, Kruse, & Georgopoulos, 1998; Romo, Hernandez, Zainos, & Salinas, 2003; Zohary, Shadlen, & Newsome, 1994). These studies report low but highly significant correlations between pairs of neurons that are recorded from the same electrode. More recent studies with multiple electrodes suggest that these correlations decay over space but remain significant even at distances of multiple millimeters (Kohn & Smith, 2005).

Simple theoretical considerations suggest that in large neural populations, average correlations should be significantly higher than in pairs of single neurons (figure 29.5*B*). The reason is that in large pools of neurons, sources of noise that are independent across the pool are averaged out while leaving the weak correlated noise unaffected; this leads to much higher correlations between the pooled responses. For example, if we assume a uniform pairwise correlation between neurons in two pools, pairwise correlations that are undetectable (e.g., $r = 10^{-3}$; solid curve, figure 29.5*B*) could lead to exceedingly high correlations between the pools for large numbers of neurons. In other words, given reasonable assumptions about the number of neurons contributing to each location in VSDI experiments, much higher correlations than are observed for pairs of single neurons are expected.

As predicted, the correlations in the VSDI signals are very high between nearby locations and fall off exponentially with space constants that are on the order of 2 mm (figure 29.5*C*). Significant correlations can be observed even at distances exceeding 4 mm.

The strong correlations at the level of the pool could contribute to the additive nature of the variability in population responses. At the level of the pool, the variance is dominated by weak correlated noise between pairs of neurons, which may be relatively stimulus-independent (but see Kohn & Smith, 2005).

Temporal Correlations in the Population Response

Temporal correlations are an important property of neural population responses with significant consequences for possible decoding mechanisms. Figure 29.5*D* shows the Pearson correlation between the amplitude of the VSDI signals in two frames as a function of their separation in time (Chen et al., 2008). The correlations are high for short intervals and fall off exponentially with a time constant of approximately 100 ms. The temporal correlations are similar in target-present and target-absent trials, consistent with the additive nature of the variability in VSDI responses. The additive variability and long-lasting temporal correlations are consistent with findings from VSDI experiments in the visual cortex of anesthetized cat (Arieli et al., 1996). Significant temporal correlations have been observed in single-unit recordings from primate visual cortex (Osborne, Bialek, & Lisberger, 2004; Uka & DeAngelis, 2003).

There are more subtle questions regarding the nature of the spatiotemporal correlations that we have not discussed here and should be addressed by future research. For example, are there higher-order correlations at the level of pools of neurons? In the retina, the observed correlations can be explained remarkably well if we assume only pairwise correlations between neighboring retinal ganglion cells

(Schneidman, Berry, Segev, & Bialek, 2006; Shlens et al., 2006). Whether this also holds in the cortex remains to be seen. Second, it is important to determine the effect of the stimulus, the behavioral state of the subject, and the specific demands of the task on the structure of the correlations. Our preliminary results suggest that at the level of large pools of neurons, the spatiotemporal correlations and the magnitude of the variability are relatively constant. More work using different complementary techniques is necessary to fully address this issue.

Exploring possible decoding mechanisms

Earlier, we presented the rationale for using the ideal Bayesian observer as a framework for exploring possible decoding mechanisms. Here, we demonstrate this approach using our own VSDI study in V1 of monkeys that were trained to perform a reaction-time detection task near the perceptual threshold (figure 29.6) (Chen et al., 2006, 2008).

Our goal was to derive the ideal Bayesian decoder for detecting the target from the monkey's VSDI signals and to compare its performance with that of several suboptimal decoders and with the performance of the monkey. For simplicity, we first describe the optimal strategy for spatial decoding, ignoring the temporal dimension by averaging the VSDI signals over a short temporal interval. We then describe the optimal strategy for decoding population responses over time, ignoring the spatial dimension by averaging the responses over a small region. This approach is reasonable because in this task the spatial responses are largely independent of time and the temporal responses are largely independent of space (e.g., figure 29.3*C*).

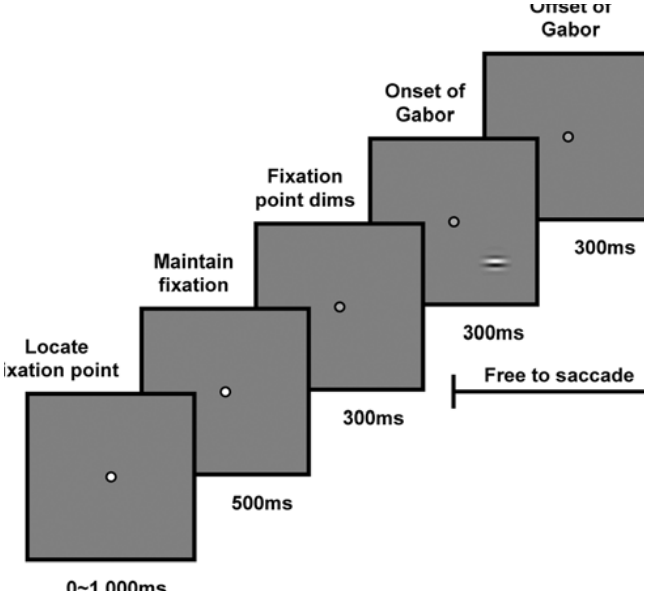


Figure 29.6 Visual detection task. Monkeys were required to detect a low-contrast Gabor patch that appeared at a known location in half of the trials. The target appeared 300 ms after the dimming of the fixation point. The monkey indicated detection by making a saccadic eye movement to the target location when it was detected but no later than 600 ms after target onset. The monkey indicated target absence by maintaining fixation for 1.5 s after fixation point dimming.

Spatial Decoding of Neural Population Responses

What is the form of an ideal Bayesian observer that performs the detection task using only the VSDI signals from the monkey's V1? We have already characterized the stimulus-evoked response (figure 29.3) and the variability (figure 29.5). Given that the spatial profiles of the stimulus-evoked response at different contrasts are scaled versions of each other and that the variability is additive Gaussian noise with significant spatial correlations, the ideal observer should use a linear summation rule of the form

$$x_{pooled} = \sum_{i=1}^{n} w_i x_i \tag{1}$$

where w_i is the weight given to response x_i from site i (Chen et al., 2006; Duda, Hart, & Stork, 2001). This pooled response is the *decision variable* that is used to determine whether the target is present or absent on a given trial. The optimal set of weights, $\mathbf{w} = \langle w_1, \ldots, w_n \rangle$, is given by

$$\mathbf{w} = \Sigma^{-1}\, \mathbf{s} \tag{2}$$

where Σ^{-1} is the inverse of the response covariance matrix Σ and $\boldsymbol{s}$ is the mean difference in response between the target-present and target-absent trials (Chen et al., 2006; Duda et al., 2001).

An equivalent way to obtain the optimal set of weights is to derive a *whitening* or *decorrelation* spatial filter that, when convolved with the population response, produces variability that is independent over space. Figure 29.7*A* shows a one-dimensional slice through the whitening filter matched to the properties of the spatial correlations in the VSDI signals (figure 29.5*C*). This filter has a sharp positive peak and a small negative trough. By applying this filter to the fitted response profile (figure 29.3*A*), we obtained the linear weights used by the optimal spatial decoder (figure 29.7*B*) (Chen et al., 2006).

The optimal weights contain a central positive region and a larger negative surround. The reason these weights have a center-surround structure is that the spatial correlations fall off more slowly over space than the signal does. Because variability in the surround, where stimulus-evoked signals are weak or absent, is still highly correlated with variability in the center, the optimal strategy is to estimate the common noise from the surround and subtract it from the center.

The detection sensitivity of the optimal decoder can be determined by measuring its performance in the detection task (Chen et al., 2006). We can also evaluate the detection

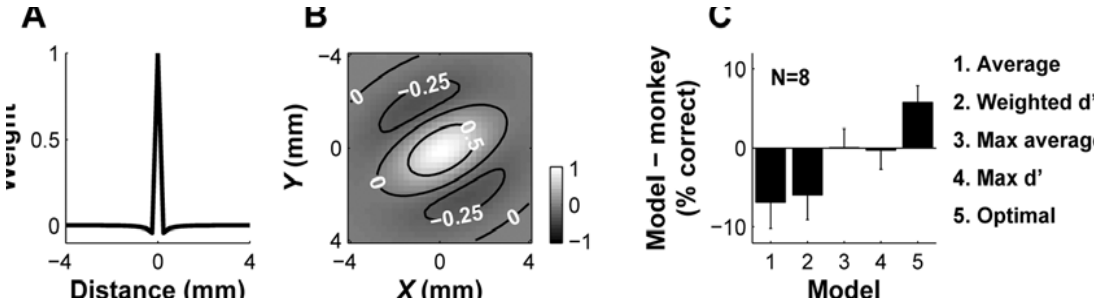


FIGURE 29.7 Optimal spatial pooling of VSDI responses in a detection task. (*A*) A one-dimensional cut through a two-dimensional spatial decorrelation (whitening) filter that removes the spatial correlations in the population responses (figure 29.5*C*). (*B*) Optimal weights for pooling the population responses over space. (*C*) Average difference in percent correct between the performance of the optimal and four suboptimal spatial pooling rules and the performance of the monkey in eight VSDI experiments.

sensitivity of other previously proposed spatial pooling rules; for example, a rule that gives equal weight to all locations (average rule, analogous to the rule used by Shadlen, Britten, Newsome, and Movshon (1996)) or a rule that weights each location based on its sensitivity (weighted d', analogous to the rule used by Geisler and Albrecht (1997)). Similarly, we can evaluate pooling rules that consider only a small region such as the location with the peak average response or the location with the highest sensitivity (maximal d').

Figure 29.7*C* shows the average difference in performance of the optimal and four suboptimal pooling rules from the performance of the monkey in eight VSDI experiments. This figure shows two surprising results. First, the optimal rule does significantly better than the monkey, demonstrating the sensitivity of the VSDI technique and showing that there are more signals in V1 than the monkey uses. Second, the two rules that pool over a large area with positive weights (average and weighted d') perform significantly worse than the monkey, while the two rules that consider only a single location perform comparably to the monkey.

The rules that pool over a large area with positive weights perform poorly, owing to the spatial correlations (figure 29.5*C*). Because the pool contains both highly sensitive and weakly sensitive neurons, averaging these together reduces signal without reducing the correlated noise. Thus when the noise is highly correlated, pooling over a small area may be better than pooling with positive weights over a larger area, even if the larger area contains signals. The only way to improve performance beyond the performance of a rule such as maximal d' is to use negative weights to cancel some of the noise. Importantly, rules that rely on a single-site perform poorly if the site is significantly smaller than 0.25×0.25 mm^2. With smaller sites, independent noise dominates the response, leading to reduced performance.

Temporal Decoding of Neural Population Responses

Next, consider the optimal Bayesian temporal decoder for detecting the target from V1 population responses in a reaction time task. Here we ignore the spatial dimension and simply consider the time course in a 1×1 mm^2 region centered on the most sensitive location. This optimal decoder evaluates V1 responses and decides, on a moment-by-moment basis, whether and when sufficient evidence that the target is present has accumulated.

To optimally decode neural population responses over time, temporal correlations in the population responses must first be removed. Analogous to space, temporal correlations can be removed by a decorrelation filter that, when convolved with the responses in single trials, produces responses that are independent across frames. To be biologically plausible, however, this filter must be causal; that is, the output of the filter at time t must depend only on the response up to time t. The whitening filter is shown in figure 29.8*A*; it has a sharp positive peak, immediately followed by a smaller and slightly longer-lasting negative peak. Such a filter could be implemented biologically with rapid excitation followed by time-lagged inhibition. The whitening operation emphasizes the response onset (and offset) relative to the sustained response (figure 29.8*B*). In other words, there is more information per unit time in the initial rising edge of the response than in the sustained response. This occurs because the response onset contains high temporal frequencies and most of the power in the correlated noise is in the low temporal frequencies.

The optimal temporal decoder takes the whitened VSDI signal in single trials and computes the dynamic posterior probability of each possible stimulus, given the observed responses (Chen et al., 2008). It then reports "target present" if the posterior probability for target presence exceeds a fixed criterion (the horizontal line in figure 29.8*C*) that is selected to maximize accuracy. The optimal temporal pooling model performed more accurately than the monkey (figure 29.8*D*). In addition, the "reaction times" of the optimal temporal pooling model (the time at which the posterior probability for target presence reached the criterion) were much faster, on average, than the monkey's reaction times (figure 29.8*E*). These results indicate that population responses provide reliable information that could guide behavior even in brief

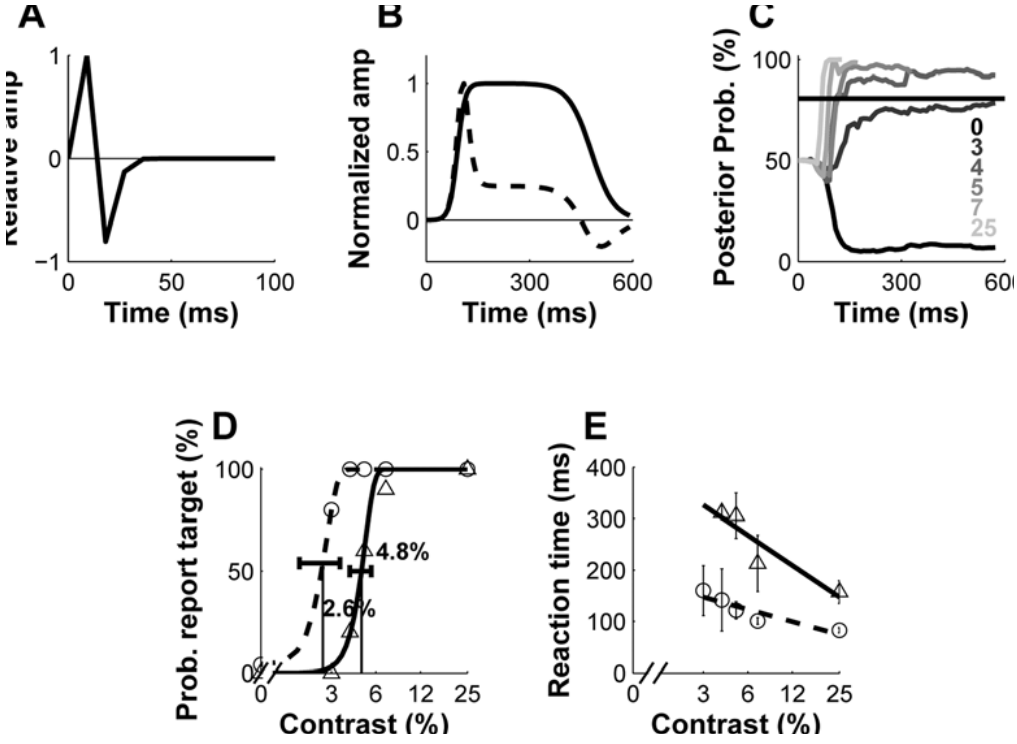


FIGURE 29.8 Optimal temporal pooling of neural population responses in a reaction-time detection task. (*A*) A causal temporal whitening filter that removes the temporal correlations in the population responses (figure 29.5*D*). (*B*) Normalized fitted time course of the response at 5% target contrast in one VSDI experiment before (solid curve) and after (dashed curve) whitening using the filter in *A*. (*C*) Posterior probability of target presence as a function of time averaged across all trials and plotted separately for each possible target contrast, including contrast zero (target absent). The horizontal black line indicates the position of the criterion on the posterior probability. (*D*) Fraction of trials in which the observer (monkey: triangles, ideal temporal decoder: circles) reported that the target was present as a function of target contrast in the same experiment as in panel *C*. The curves indicate best-fit Weibull functions. The numbers are the stimulus contrasts required for an accuracy of 75% correct. (*E*) Average and standard deviation of reaction time of monkey (triangles) and ideal observer (circles) as a function of target contrast. The reaction time of the ideal observer was taken as the time the posterior probability for target presence exceeded the threshold.

temporal intervals (~100 ms). The mean and the variance of both the ideal observer's and the monkey's reaction times increase with decreasing target contrast, but at a faster rate for the monkey than for the ideal observer (figure 29.8*E*).

As with spatial pooling, one advantage of deriving the optimal decoder is that it can serve as a benchmark to which suboptimal models can be compared. We evaluated the performance of a simple model in which the VSDI responses are summed until a fixed threshold is reached. This model performs significantly worse than the monkey. We also evaluated an optimally shaped "running integrator" model that integrates the whitened responses over a window of about 100 ms. Because most of the information in our task was concentrated in the rising edge of the response and because the temporal profiles of the response at different contrasts differ only in latency and amplitude, this running integrator model performed almost as well as the ideal observer.

Note that the ideal observer would have performed much better had the responses been statistically independent over time, demonstrating that in our task, the detrimental effect of temporal correlations cannot be entirely overcome by optimal pooling. Finally, note that the performance of the optimal temporal decoder (and running integrator decoder) can be improved further by combining signals over space using the optimal spatial pooling rule rather than averaging the signals in a 1.0-mm^2 region (Chen et al., 2008).

Discussion

This chapter began with two central claims that are relevant to the goal of understanding encoding and decoding by neural populations in the mammalian cortex. First, emergent properties in large neural populations make it essential to augment single-neuron recording with techniques that measure the responses of large populations of neurons simultaneously. Second, in formulating and testing hypotheses for population encoding/decoding, it can be highly beneficial to derive and evaluate optimal encoding and decoding strategies. To illustrate the first claim, we reviewed some emergent properties that have been observed in V1: widespread

responses even from maximally localized stimuli, highly sparse representations with most of the population response arising from weakly responding neurons, rapid response dynamics over large areas of cortex, additive but not multiplicative population noise, and large spatial and temporal noise correlations. To illustrate the second claim, we described the optimal decoding strategy for VSDI signals recorded from behaving monkeys in a reaction-time detection task and how this optimal decoder provides insight into specific questions about neural decoding.

Although progress is being made, the rigorous study of population responses in sensory and motor areas of the cortex is just beginning; indeed, the results obtained to date raise more questions than they answer. Next we discuss some of the relevant issues.

Weakly Versus Strongly Responding Neurons An important emergent property of large population responses in V1 is the dominance of relatively weakly responding neurons in the total population activity (figure 29.4). An open question is whether the weakly responding neurons are ignored or used by subsequent decoding mechanisms. It is not uncommon from the perspective of single-neuron electrophysiology to assume that those neurons that are most sensitive to a stimulus are the ones that carry most of the information used by the brain, but this need not be the case. For example, magnocellular neurons in the LGN are much more sensitive to contrast than are parvocellular neurons, and hence one might expect them to dominate performance in contrast detection, but in fact, the much more numerous but weaker responding parvocellular neurons dominate in most contrast detection tasks (Merigan, Katz, & Maunsell, 1991).

To further illustrate the potential significance of weakly responding neurons, consider the study of choice probability (trial-by-trial correlations between neural and behavioral variability) at the single-neuron level. If a large pool of weakly responding neurons were contributing as much to a subject's choice as a small pool of strongly responding neurons, recording from a single strongly responding neuron could easily yield a measurable correlation with behavior, whereas recording from a single weakly responding neuron could easily yield no measurable correlation. Obviously, it would be a mistake to interpret the lack of correlation in the weakly responding neuron as evidence against a major role for the weakly responding neurons in the subject's choice.

These considerations provide an additional illustration of a central theme of this chapter: that effects that are very weak at the single-neuron level could have a dominant role at the level of the pool. Therefore a general conclusion is that one should be cautious when making predictions based on single-unit measurements.

Sources of Correlated Noise Another important emergent property in V1 is the widespread and large spatial and temporal correlations in the variability of the population response. These correlations can have profound consequences for decoding, and they are entirely expected when a large number of neural inputs, having weakly correlated noise, are summed. Thus, it is important to identify and characterize the sources responsible for the small correlated noise that is shared between the neurons in a population. For example, it is possible that weak correlated noise must always be present, owing to the inevitable sharing of inputs between neurons. If so, then every time a large amount of convergence is required in a neural circuit, there will be the need for a decorrelating mechanism that can cancel most of the correlated noise.

Decoding a Neural Population Response with a Neural Population In the description of optimal and suboptimal candidate decoders, we were not explicit about how they might be implemented. In all likelihood, the decoding of population responses is implemented with another neural population. In fact, it is likely that, at every step along a sensorimotor pathway (from sensory encoding, to decision computation, all the way to the activation of muscle fibers), the stimulus and/or motor response is represented by the activity of a large neural population, because that is the obvious way to obtain robust behavior without ever requiring any specific neuron to be as robust as the behavior.

Decoding Population Responses in Different Cortical Areas Given the similarities in neural anatomy across the cortex, it is quite possible that in all sensory and motor areas (as in V1), even the most localized inputs are encoded by population activity that extends over at least several square millimeters. However, there may also be some substantial differences in the properties of population responses across areas. For example, it is possible that early sensory areas contain a more sparse representation than higher sensory areas because they must represent many stimulus dimensions within the same area. This could have important consequences for the properties of population responses and hence for subsequent decoding.

Decoding Population Responses in Different Tasks The ideal Bayesian spatial decoder developed for our detection task (figure 29.7*B*) can be extended to other detection and discrimination tasks. As long as the variability is consistent with an additive Gaussian noise, the optimal weights are given by equation 2. Because the variability is additive, the only factor that determines the task-dependent component of the optimal weights is the difference in the mean response between the two stimulus conditions, $\boldsymbol{s}$.

Recall that the correlated noise is dominated by low spatial frequencies (figure 29.5*C*); therefore the whitening operation of the ideal observer amounts to high-pass filtering of $\boldsymbol{s}$. Figure 29.9 illustrates the expected impact of this whitening operation on the optimal weights for three detection tasks and one discrimination task. The upper row in figure 29.9 shows the stimuli; the middle row shows $\boldsymbol{s}$, the difference in the mean hypothetical responses for each task; and the bottom row the optimal weights. In the detection tasks (figures 29.9*A*–29.9*C*), the difference in response to the stimulus and blank is dominated by low spatial frequencies; therefore the optimal weights are strongly affected by the whitening operation. On the other hand, in the discrimination task (figure 29.9*D*), the difference in response to the two oriented stimuli is dominated by high spatial frequencies (i.e., comparable to the spatial frequency of orientation columns in macaque V1). In this case, the whitening operation has little impact on the optimal weights because the subtraction in $\boldsymbol{s}$, by itself, is sufficient to remove the long-range spatial correlations. However, even in this case, temporal whitening with time lagged inhibition would still be beneficial for minimizing the impact of the temporal correlation.

Feedback Within and Across Populations The role of feedback in encoding and decoding by populations of neurons is still largely unknown. Feedback can play an important role in optimizing encoding of sensory stimuli for the specific demands of the task (e.g., Li, Piech, & Gilbert, 2004). Feedback can also play an active role in the decoding process through dynamic interactions between the encoding and decoding stages. One potential signature of feedback effects could be delayed modulations of the stimulus-evoked responses and/or the response variability. In our reaction-time detection task, however, both the evoked response and the response variability were relatively constant throughout stimulus presentation, suggesting that feedback does not play a strong role in this task.

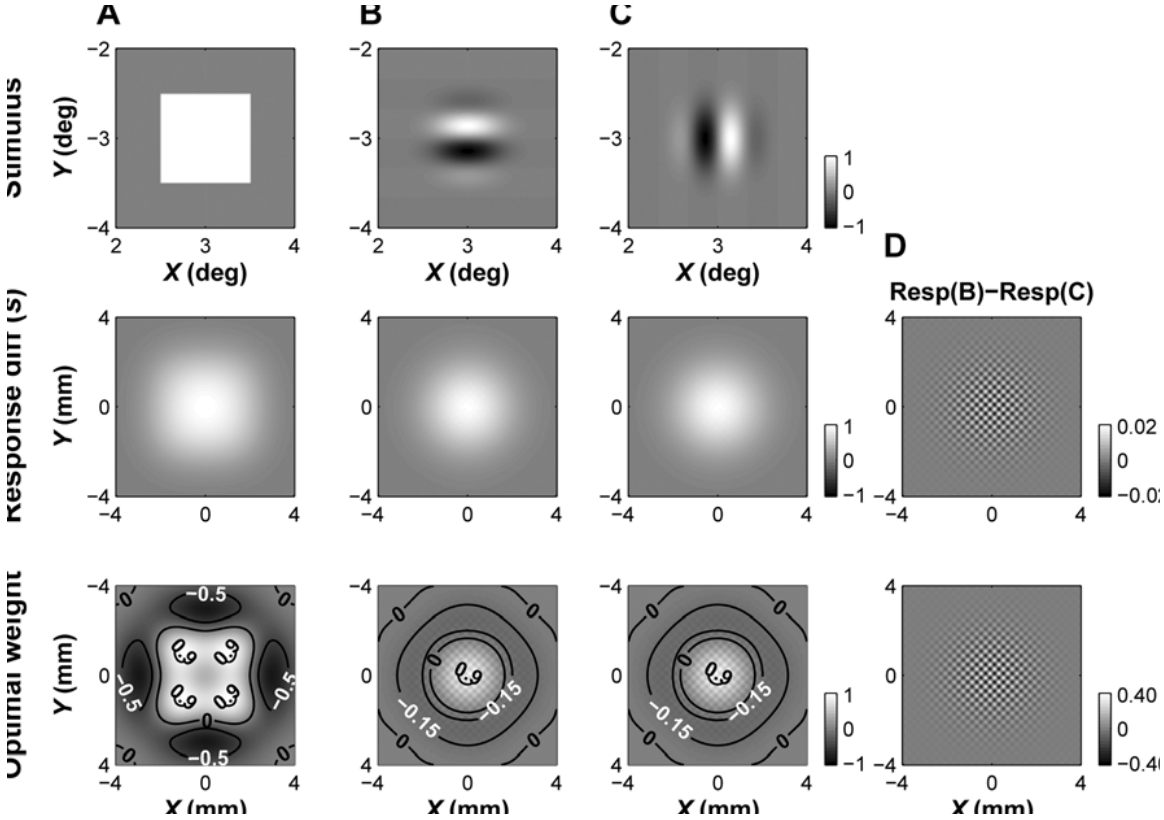


Figure 29.9 Optimal linear weights in three different perceptual tasks based on hypothetical population responses. (*A*) Detection of a low-contrast square. (*B*) Detection of a low-contrast horizontal Gabor patch. (*C*) Detection of a low-contrast vertical Gabor patch. (*D*) Discrimination between the Gabor patches in panels *B* and *C*. (*A*–*C*) top panel: stimulus; middle panel: hypothetical response in an 8×8 mm^2 patch of cortex; bottom panel: optimal linear weights obtained by applying the whitening filter in figure 29.7*A* to the population responses in the middle panel. (*D*) top panel: response to the horizontal Gabor in panel *B* minus the response to the vertical Gabor in panel *C*. The orientation-selective response was modeled as a high-spatial-frequency activation pattern (2.5 cycles/mm) with amplitude that is 10% of the amplitude of the Gaussian envelope of the population response. (*D*) bottom panel: optimal linear weights for discriminating between the Gabor patches at the two orientations.

Identifying the Actual Neural Code and Decoder The goal of an ideal observer analysis is to determine how population responses *should* be pooled over space and time to perform a specific task optimally. As was discussed before, this approach can, in principle, be used to reject possible combinations of codes and decoders if their sensitivity falls significantly short of the subject's sensitivity. The fact that the ideal observer does better than the monkeys in our detection task shows that in this task, there is no need to assume an actual code with a finer spatial and/or temporal resolutions than the one provided by VSDI. An important goal for future research is to determine whether this holds in other tasks. For example, it is possible that in an orientation discrimination task, an ideal observer using VSDI signals from V1 would perform significantly worse than the subject, owing to the coarse spatial pooling that is inherent to this technique.

The finding that an ideal observer performs significantly better than the subject does not necessarily imply that the subject is using a suboptimal decoding strategy. Population responses could be pooled by using the optimal pooling strategy but be degraded by subsequent sources of noise. For a more complete discussion of why the monkeys might perform suboptimally in our detection task, see Chen and colleagues (2006, 2008).

Ultimately, the goal of this line of research is to determine what are the *actual code* and *actual decoder* used by the observer. We are a long way from being able to address this question even in the simplest perceptual and motor tasks. Next, we briefly mention two approaches that could be used to address these questions.

One potential approach is to examine the trial-by-trial covariations between neural and behavioral responses (choice probability). Previous studies of neural and behavioral performances near psychophysical threshold demonstrated weak but significant covariation between the activity of single neurons and behavioral responses (e.g., Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Cook & Maunsell, 2002; Palmer et al., 2007; Purushothaman & Bradley, 2005). If such correlations can be measured at the population level, their nature could provide useful information regarding the decoding mechanisms used by the subject.

A second approach that could be used to study the actual decoder is to perturb population responses in specific ways that are designed to distinguish between possible decoding mechanisms (e.g., Lee, Rohrer, & Sparks, 1988). Combining careful behavioral and neurophysiological measurements with better techniques for selectively perturbing brain activity at the population level (e.g., genetic-based techniques: Tan et al., 2006; Zhang et al., 2007) is a promising direction for testing candidate decoding models.

Conclusions

This chapter reviewed recent progress in understanding neural population coding in the mammalian cortex. This research area is clearly in its infancy. Making further progress will necessitate improving existing techniques, as well as developing new techniques, for monitoring and manipulating neural population responses in behaving subjects. It is unlikely that any single technique will provide access to the real-time activity of all the neurons in a given cortical area that could potentially contribute to behavior. Therefore it is important to develop a quantitative understanding of the relationships between the measurements of neural populations obtained with different techniques at different spatial scales. Analyzing simultaneous measurements of population responses and behavioral performance, within a Bayesian ideal-observer framework, is a powerful approach for addressing fundamental issues of population encoding and decoding.

acknowledgments We thank W. Bosking, C. Michelson, C. Palmer, and Z. Yang for discussions, and T. Cakic for technical support. This work was supported by National Eye Institute Grants EY-016454 and EY-016752 to E. Seidemann and EY-02688 to W. S. Geisler and by a Sloan Foundation Fellowship to E. Seidemann.

NOTE

1. Note that this measure of information is related to, but differs from, traditional measures in information theory (Cover & Thomas, 2006). For example, Shannon information (mutual information) is appropriate for characterizing the potential bit rate of information transfer through a noisy channel when the goal is input reconstruction, but mutual information is not monotonically related to the trial-by-trial accuracy of an ideal observer in a discrimination or classification task (Geisler, Albrecht, Salvi, & Saunders, 1991). Fisher information can be monotonically related to the performance of the ideal observer, although the ideal observer provides the more general measure.

REFERENCES

Abbott, L. F., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Comput.*, *11*(1), 91–101.

Albrecht, D. G., Geisler, W. S., Frazor, R. A., & Crane, A. M. (2002). Visual cortex neurons of monkeys and cats: Temporal dynamics of the contrast response function. *J. Neurophysiol.*, *88*(2), 888–913.

Arieli, A., Grinvald, A., & Slovin, H. (2002). Dural substitute for long-term imaging of cortical activity in behaving monkeys and its clinical implications. *J. Neurosci. Methods*, *114*(2), 119–133.

Arieli, A., Sterkin, A., Grinvald, A., & Aertsen, A. (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science*, *273*(5283), 1868–1871.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.*, *7*(5), 358–366.

Bair, W., Zohary, E., & Newsome, W. T. (2001). Correlated firing in macaque visual area MT: Time scales and relationship to behavior. *J. Neurosci.*, *21*(5), 1676–1697.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.*, *13*(1), 87–100.

Chen, Y., Geisler, W. S., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nat. Neurosci.*, *9*(11), 1412–1420.

Chen, Y., Geisler, W. S., & Seidemann, E. (2008). Optimal temporal decoding of V1 population responses in a reaction-time detection task. *J. Neurophysiol.*, *99*(3), 1366–1379.

Churchland, M. M., Yu, B. M., Sahani, M., & Shenoy, K. V. (2007). Techniques for extracting single-trial activity patterns from large-scale neural recordings. *Curr. Opin. Neurobiol.*, *17*(5), 609–618.

Cook, E. P., & Maunsell, J. H. R. (2002). Dynamics of neuronal responses in macaque MT and VIP during motion detection. *Nat. Neurosci.*, *5*(10), 985–994.

Cover, T., & Thomas, J. (2006). *Elements of information theory* (2nd ed.). New York: Wiley.

de Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vis. Res.*, *22*(5), 545–559.

de Valois, R. L., & de Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.

DeAngelis, G. C., Ghose, G. M., Ohzawa, I., & Freeman, R. D. (1999). Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons. *J. Neurosci.*, *19*(10), 4046–4064.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: John Wiley.

Gawne, T. J., & Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical-neurons. *J. Neurosci.*, *13*(7), 2758–2771.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.*, *96*(2), 267–314.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.*, *59*(1), 167–192.

Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination, and identification. *Vis. Neurosci.*, *14*(5), 897–919.

Geisler, W. S., Albrecht, D. G., Salvi, R. J., & Saunders, S. S. (1991). Discrimination performance of single neurons: Rate and temporal-pattern information. *J. Neurophysiol.*, *66*(1), 334–362.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley.

Grinvald, A., & Hildesheim, R. (2004). VSDI: A new era in functional imaging of cortical dynamics. *Nat. Rev. Neurosci.*, *5*(11), 874–885.

Grinvald, A., Shoham, D., Shmuel, A., Glaser, D. E., Vanzetta, I., Shtoyerman, E., et al. (1999). In-vivo optical imaging of cortical architecture and dynamics. In U. Windhorst & H. Johansson (Eds.), *Modern techniques in neuroscience research* (pp. 893–969). New York: Springer.

Hubel, D. H., & Weisel, T. N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.*, *198*(1130), 1–59.

Johnson, K. O. (1980). Sensory discrimination: Neural processes preceding discrimination decision. *J. Neurophysiol.*, *43*(6), 1793–1815.

Kelly, R. C., Smith, M. A., Samonds, J. M., Kohn, A., Bonds, A. B., Movshon, J. A., & Lee, T. S. (2007). Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex. *J. Neurosci.*, *27*(2), 261–264.

Kerr, J. N. D., de Kock, C. P. J., Greenberg, D. S., Bruno, R. M., Sakmann, B., & Helmchen, F. (2007). Spatial organization of neuronal population responses in layer 2/3 of rat barrel cortex. *J. Neurosci.*, *27*(48), 13316–13328.

Kerr, J. N. D., & Denk, W. (2008). Imaging in vivo: Watching the brain in action. *Nat. Rev. Neurosci.*, *9*(3), 195–205.

Kohn, A., & Smith, M. A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.*, *25*(14), 3661–3673.

Lee, C. K., Rohrer, W. H., & Sparks, D. L. (1988). Population coding of saccadic eye-movements by neurons in the superior colliculus. *Nature*, *332*(6162), 357–360.

Lee, D., Port, N. L., Kruse, W., & Georgopoulos, A. P. (1998). Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *J. Neurosci.*, *18*(3), 1161–1170.

Li, W., Piech, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.*, *7*(6), 651–657.

McIlwain, J. T. (1986). Point images in the visual system: New interest in an old idea. *Trends Neurosci.*, *9*, 354–358.

Merigan, W. H., Katz, L. M., & Maunsell, J. H. (1991). The effects of parvocellular lateral geniculate lesions on the acuity and contrast sensitivity of macaque monkeys. *J. Neurosci.*, *11*(4), 994–1001.

Nauhaus, I., Benucci, A., Carandini, M., & Ringach, D. (2008). Neuronal selectivity and local map structure in visual cortex. *Neuron*, *57*(5), 673–679.

Nicolelis, M. A. L., & Ribeiro, S. (2002). Multielectrode recordings: The next steps. *Curr. Opin. Neurobiol.*, *12*(5), 602–606.

Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2004). Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *J. Neurosci.*, *24*(9), 2065–2076.

Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P., & Reid, R. C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, *433*(7026), 597–603.

Osborne, L. C., Bialek, W., & Lisberger, S. G. (2004). Time course of information about motion direction in visual area MT of macaque monkeys. *J. Neurosci.*, *24*(13), 3210–3222.

Palmer, C., Cheng, S.-Y., & Seidemann, E. (2007). Linking neuronal and behavioral performance in a reaction-time visual detection task. *J. Neurosci.*, *27*(30), 8122–8137.

Priebe, N. J., & Ferster, D. (2008). Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. *Neuron*, *57*(4), 482–497.

Purushothaman, G., & Bradley, D. C. (2005). Neural population code for fine perceptual decisions in area MT. *Nat. Neurosci.*, *8*(1), 99–106.

Romo, R., Hernandez, A., Zainos, A., & Salinas, E. (2003). Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron*, *38*(4), 649–657.

Schneidman, E., Berry, M. J., Segev, R., & Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, *440*(7087), 1007–1012.

Seidemann, E., Arieli, A., Grinvald, A., & Slovin, H. (2002). Dynamics of depolarization and hyperpolarization in the frontal cortex and saccade goal. *Science*, *295*(5556), 862–865.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.*, *16*(4), 1486–1510.

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., et al. (2006). The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.*, *26*(32), 8254–8266.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.*, *24*, 1193–1216.

Slovin, H., Arieli, A., Hildesheim, R., & Grinvald, A. (2002). Long-term voltage-sensitive dye imaging reveals cortical dynamics in behaving monkeys. *J. Neurophysiol.*, *88*(6), 3421–3438.

Snippe, H. P., & Koenderink, J. J. (1992). Information in channel-coded systems: Correlated receivers. *Biol. Cybern.*, *67*(2), 183–190.

Sompolinsky, H., Yoon, H., Kang, K. J., & Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Phys. Rev. E*, *64*(5), 051904-1–051904-11.

Stosiek, C., Garaschuk, O., Holthoff, K., & Konnerth, A. (2003). In vivo two-photon calcium imaging of neuronal networks. *Proc. Natl. Acad. Sci. USA*, *100*(12), 7319–7324.

Tan, E. M., Yamaguchi, Y., Horwitz, G. D., Gosgnach, S., Lein, E. S., Goulding, M., et al. (2006). Selective and quickly reversible inactivation of mammalian neurons in vivo using the drosophila allatostatin receptor. *Neuron*, *51*(2), 157–170.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual-cortex. *Vis. Res.*, *23*(8), 775–785.

Tootell, R. B. H., Switkes, E., Silverman, M. S., & Hamilton, S. L. (1988). Functional-anatomy of macaque striate cortex: 2. Retinotopic organization. *J. Neurosci.*, *8*(5), 1531–1568.

Uka, T., & DeAngelis, G. C. (2003). Contribution of middle temporal area to coarse depth discrimination: Comparison of neuronal and psychophysical sensitivity. *J. Neurosci.*, *23*(8), 3515–3530.

Van Essen, D. C., Newsome, W. T., & Maunsell, J. H. R. (1984). The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. *Vis. Res.*, *24*(5), 429–448.

Yang, Z., Heeger, D. J., & Seidemann, E. (2007). Rapid and precise retinotopic mapping of the visual cortex obtained by voltage sensitive dye imaging in the behaving monkey. *J. Neurophysiol.*, *98*(2), 1002–1014.

Zhang, F., Wang, L.-P., Brauner, M., Liewald, J. F., Kay, K., Watzke, N., et al. (2007). Multimodal fast optical interrogation of neural circuitry. *Nature*, *446*(7136), 633–639.

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*(6485), 140–143.

30 Perceptual Filling-in: From Experimental Data to Neural Network Modeling

RAINER GOEBEL AND PETER DE WEERD

ABSTRACT Recognizing objects, performing goal-directed actions, and navigating the environment are capabilities that crucially depend on the ability to correctly segregate and perceive surfaces. Surface perception results from computations that involve multiple processing levels in the visual system. The nature of these computations has been a matter of persistent debate. The contribution of this chapter to this debate is threefold. In the first part ("Reconstructive Processes Contributing to Surface Perception"), an overview is given of empirical studies that inform and constrain computational models of surface reconstruction. In the second part ("A Computational Model for Modal Texture Filling-in"), empirically supported principles of surface perception and known architecture of early visual areas are used to build a computational model of neural activity corresponding to visual filling-in of surface texture in early visual areas. The model explicitly simulates subthreshold and suprathreshold activity and therefore generates predictions not only for the activity distributions of spiking neurons, but also for functional magnetic resonance imaging (fMRI). The third part, "Insights, Limitations, and Future Research Directions," provides a discussion of the main insights following from the modeling results. The model demonstrates the consequences of subthreshold neural spread of the BOLD signal for the activity distribution obtained with stimuli typically used to investigate perceptual filling-in, and it provides insight into the divergent data from human fMRI and neurophysiological experiments in animals. The model's architecture, in which surface-related activity in low-level regions is validated by recurrent loops involving higher-order areas, is in line with theories of conscious perception. Limitations of the model will be discussed, and future research directions will be proposed.

RAINER GOEBEL Faculty of Psychology and Neuroscience, Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands; Netherlands Institute for Neuroscience (NIN), an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), Amsterdam, The Netherlands
PETER DE WEERD Faculty of Psychology and Neuroscience, Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands

Reconstructive processes contributing to surface perception

The Problem of Surface Perception

A brief theoretical background The question how surfaces are represented follows from the normalization of the visual image by antagonistic filters in retina and LGN (Grossberg, 2003a). Normalization leads to a strong emphasis on discontinuities in the light distribution on the retina and a severe loss of signal related to regions of homogenous stimulation. Although that loss is not complete (see the section entitled "Features of Early Visual System Compatible with Active Interpolation"), it raises the question how the vivid experience of surfaces arises in the visual system. This has remained a matter of intense debate (Pessoa & Neumann, 1998; Pessoa, Thompson, & Noë, 1998; Komatsu, 2006), and three classes of mechanisms can be distinguished.

According to *active interpolation theory*, surface reconstruction involves active interpolation of surface features from a surface's edges inward. The neural substrate of surface-related spreading activation is hypothesized to exist in early visual areas that are retinotopically organized. In this view, boundary representations initiate and contain the spreading of surface information (Walls, 1954; Gerrits, de Haan, & Vendrik, 1966; Gerrits & Vendrik, 1970; Grossberg, 2003a). First, contour interpolation mechanisms use local discontinuities to construct boundary representations. Second, surface interpolation mechanisms use measurements close to the contours as seeding points for the spread of surface feature, and inhibitory signals emanating from boundary representations contain spreading activity within a surface's cortical projection. Computational models that incorporate an interaction between boundary and surface-related mechanisms perform well in explaining a wide range of visual illusions (Grimson, 1982; Grossberg, 1987a, 1987b, 1997, 2003a, 2003b; Todorovic, 1987; Grossberg & Todorovic, 1988; Arrington, 1994; Gove, Grossberg, & Mingolla, 1995; Grossberg & Raizada, 2000; Raizada & Grossberg, 2001).

Second, *multiscale spatial frequency filtering theory* suggests that surface representations can be derived directly from low-spatial-frequency information (McCourt, 1982; Stromeyer et al., 1984; Purves, Shimpi, & Lotto, 1999; Blakeslee, Pasieka, & McCourt, 2005, Blakeslee & McCourt, 2008; Dakin & Bex, 2003). This theory is compatible with a correlate of surface perception in retinotopic visual areas. However, because low spatial frequencies are processed prior to higher ones, surface representations might emerge before the completion of precise boundary representations, which require processing of high spatial frequencies (Hughes, Nozawa, & Kitterle, 1996).

Third, and in contrast to the previous two proposals, Denett (1991) proposed a *symbolic encoding theory* in which surface qualities are not encoded in retinotopic areas but are encoded implicitly (symbolically) as a lack of discontinuities in between identified surface boundaries. No activity is expected in lower visual areas corresponding to the perceived aspects of surfaces, and surface encoding might take place entirely at a higher, nonretinotopic level in the visual system, where surface representations might be an integral aspect of object representations (Desimone & Ungerleider, 1989; Wang, Tanaka, & Tanifuji, 1996; Biederman, 2000; Haxby et al., 2001; Kayaert, Biederman, & Vogels, 2003).

We take the viewpoint that the perceptual quality of surfaces is at least partly reconstructed from local cues in the retinal image, and in this first part of the chapter, we review evidence that is directly relevant for the active interpolation theory. However, evidence for active interpolation does not exclude contributions of low-spatial-frequency filtering and of feedback from high-level areas.

Experimental paradigms Two types of paradigm have been used to investigate whether active interpolation mechanisms exist (for a review, see Martinez-Conde, 2006). One approach is to directly study the perception of surfaces under normal viewing conditions, in intact or damaged visual systems (Bender & Teuber, 1946; Sergent, 1988), and to test whether surface perception would be associated with a (relatively) fast-spreading mechanism. An alternative approach entails the use of image stabilization, which produces filling-in illusions that (despite their delay from stimulus onset) are believed to yield insight into mechanisms of normal surface perception. When the image is stabilized artificially (e.g., using contact lenses), perceptual filling-in of the stabilized images occurs within a few hundred milliseconds. This has been demonstrated for color as well as brightness (Riggs & Ratliff, 1952; Ditchburn & Ginsborg, 1962; Riggs, Ratliff, Cornsweet, & Cornsweet, 1953; Gerrits et al., 1966, Yarbus, 1967). Taking advantage of the near-perfect stabilization of entopic images, Coppola and Purves (1996) demonstrated filling-in within as little as 80 ms. By contrast, when the image is stabilized by using a strategy of maintained fixation away from a figure (Troxler, 1804), perceptual filling-in of the figure by the background can take many seconds, the length of the delay depending on the exact stimulus conditions (e.g., De Weerd et al., 1998). This delayed type of filling-in has been demonstrated for color, brightness, and (dynamic) texture (Ramachandran & Gregory, 1991; Spillmann & Kurtenbach, 1992; Ramachandran, Gregory, & Aiken, 1993; Fujita, 1993).

The different onset time of filling-in in the different paradigms fits with the active interpolation model, in which boundary construction precedes surface filling-in, so the spreading activation related to surface filling-in can be contained by inhibition from boundary representations. According to this model, in the brief periods in between saccades during which an observer's eyes rest on a specific point of interest during normal vision, surface interpolation takes place within roughly 100 ms of the completion of boundary representations (the exact time interval depending on surface size). An absence of retinal input within the confines of a cortical surface representation would not affect the cortical interpolation mechanisms, hence the quasi-instantaneous filling-in of the blind spot (Fiorani, Rosa, Gattass, & Rocha-Miranda, 1992; Komatsu & Murakami, 1994; Matsumoto & Komatsu, 2005) and retinal scotomas (Murakami, Komatsu, & Kinoshita, 1997). However, when the image is stabilized on the retina by artificial means or maintained fixation, the ensuing filling-in of a figure by the background is delayed by a time period that corresponds to the time for figure boundary mechanisms to adapt (Clarke, 1957). After a release from inhibition resulting from adaptation, surface feature from the background is interpolated into the area previously occupied by the figure (Tremere, Pinaud, & De Weerd, 2003).

Definitions of filling-in and adaptation *Perceptual filling-in* refers to the spreading of surface feature across a region in the visual field where that feature is physically absent. The term *filling-in* does not always aptly describe the percept, as in some paradigms, feature mixing can be experienced (e.g., Hsieh & Tse, 2006) rather than one region becoming filled in by another. We use the term *filling-in* with that caveat in mind. Furthermore, the term *filling-in* is used to refer to the perceptual phenomenon, and the term *interpolation* or *spreading activation* to refer to the underlying neural mechanism.

The term *adaptation* in a Troxler paradigm is likely to be related to more than the adaptation of local boundary representations and includes more global mechanisms that process the organization of visual scenes and that determine which image areas are labeled as *figure* and *background*. Nevertheless, for the purpose of computational modeling, the

concept of adaptation will be linked with the weakening of boundary representations over time, and the concept of filling-in will be linked with inward flow of background feature into the former figure region.

Features of Early Visual System Compatible with Active Interpolation

A possible anatomical basis for boundary and surface processing mechanisms in V1, V2, and V4 An essential aspect of active interpolation models of surface perception is the distinction between networks that analyze boundaries and networks that produce surfaces (Walls, 1954; Gerrits & Vendrik, 1970; Grossberg, 1987a, 1987b, 2003a). This distinction is loosely based upon *functional differences between anatomically defined compartments in V1, V2, and V4*. By using cytochrome oxidase staining, V1 can be subdivided into regions of dense staining (blobs) and less dense staining (interblobs). Blobs in V1 project to thin stripes in V2, and interblobs in V1 project to interstripes as well as thick stripes in V2 (Livingstone & Hubel, 1983; Roe & Ts'o, 1995; Sincich & Horton, 2005) (figure 30.1*A*). Interestingly, V2 thin stripes and interstripes project to separate domains in V4 (Nakamura, Gattass, Desimone, & Ungerleider, 1993; Yoshioka & Dow, 1996; Xiao, Zych, & Felleman, 1999). The data suggest the existence of two processing streams within early to midlevel visual cortex. Both arise from a mix of parvocellular (P) and magnocellular (M) thalamic inputs arriving in layers 4Cb and 4Ca of V1, respectively (Blasdel & Lund, 1983; Nealey & Maunsell, 1994). Furthermore, although feedforward connectivity is predominantly specific within the two streams, feedback (Livingstone & Hubel, 1983; Roe & Ts'o, 1999; Angelucci et al., 2002; Xiao & Felleman, 2004; Angelucci & Bressloff, 2006), as well as lateral connectivity (Gilbert & Wiesel, 1989, McGuire, Gilbert, Rivlin, & Wiesel, 1991; Lund, Yoshioka, & Levitt, 1993; Malach, Tootell, & Malonek, 1994; Yoshioka, Blasdel, Levitt, & Lund, 1996; Yabuta & Callaway, 1998b) shows a mix of specificity and cross-stream connectivity, with the best specificity in layer III and less specificity outside (Ts'o & Gilbert, 1988; Farias, Gattass, Piñon, & Ungerleider, 1997; Shipp & Zeki, 2002). Hence the hypothesized processing streams in early visual cortex are strongly interconnected, and their signals likely interact.

Because of the mixing of P and M inputs in V1, the perceptual contribution of different compartments in early visual areas cannot be inferred from the functional properties of P and M cells, and early proposals that attempted to assign discrete functions such as shape, color, and motion to different anatomical compartments have proven incorrect (for a review, see Sincich & Horton, 2005). Nevertheless, the anatomically defined compartments tend to be characterized by different functional properties. Relevant for the present review is that V1 blobs (Friedman, Zhou, & von der Heydt, 2002; Roe, Lu, & Hung, 2005) and V2 thin stripes (Roe & Ts'o, 1995; Ts'o, Roe, & Gilbert, 2001; Wang, Xiao, & Feldman, 2007) contain neurons that are well stimulated by large chromatic and achromatic homogenous surfaces overlaying their receptive fields (RFs). Other studies that did not link recording sites to anatomical compartments confirmed the presence of responses to homogenous surfaces in monkey V1 (Kayama, Riso, Bartlett, & Doty, 1979; Maguire & Baizer, 1982; Kinoshita & Komatsu, 2001; Huang & Paradiso, 2008) and in cat area 17 (Bonhoeffer, Kim, Malonek, Shoham, & Grinvald, 1995; Shoham, Hübener, Schulze, Grinvald, & Bonhoeffer, 1997). By contrast, V1 interblob and V2 interstripe regions contain neurons that show preferential responses to chromatic and achromatic oriented lines and edges (Blasdel, Lund, & Fitzpatrick, 1985; DeYoe & Van Essen, 1988; Blasdel, Obermayer, & Kiorpes, 1995; Fitzpatrick, Lund, & Blasdel, 1985; Hubel & Livingstone, 1987, 1990; Bartfeld & Grinvald, 1992; Blasdel, 1992; Yabuta & Callaway, 1998a; Landisman & Ts'o, 2002; Lu & Roe, 2008). Furthermore, V1 interblobs, as well as V2 thick stripes and interstripes contribute to the analysis of local motion and stereo cues (DeYoe & Van Essen, 1988), but the contribution of these cues to surface segregation and interpolation falls outside the scope of this chapter (Grimson, 1982; Buckley, Frisby, & Mayhew, 1989; Frisby et al., 1995; Treue, Andersen, Ando, & Hildreth, 1995; Hillis, Watt, Landy, & Banks, 2004). On the basis of anatomy and physiological data with stimuli defined by luminance and color, we suggest that there are separate (but interacting) processing streams for contour and surface processing, extending from V1 and V2 into V4.

A possible anatomical basis for spreading activation: Retinotopy and lateral connectivity Early visual cortex shows retinotopic organization (Gattass, Gross, & Sandell, 1981; Gattass, Sousa, & Gross, 1988; Sereno et al., 1995) and therefore provides a substrate for isomorphic spreading activation, while the lack of retinotopy in higher-order cortex (beyond V4) limits its contribution to isomorphic representations. In retinotopic areas, both feedback and lateral connectivity could contribute to spreading. However, psychophysical estimates of the speed of spreading activation fit better with the transmission speed of lateral connectivity (0.2 m/s or 10–30 ms/degree) than with feedback, which is an order of magnitude faster (for review, see Angelluci & Bresloff, 2006). We therefore suggest that the *speed of surface-related spreading activation is set by properties of lateral connectivity*.

Further, cortical compartments with differentiable functions are *intertwined* in each retinotopic map, often at a submillimeter scale (figure 30.1*A*). This implies that functional

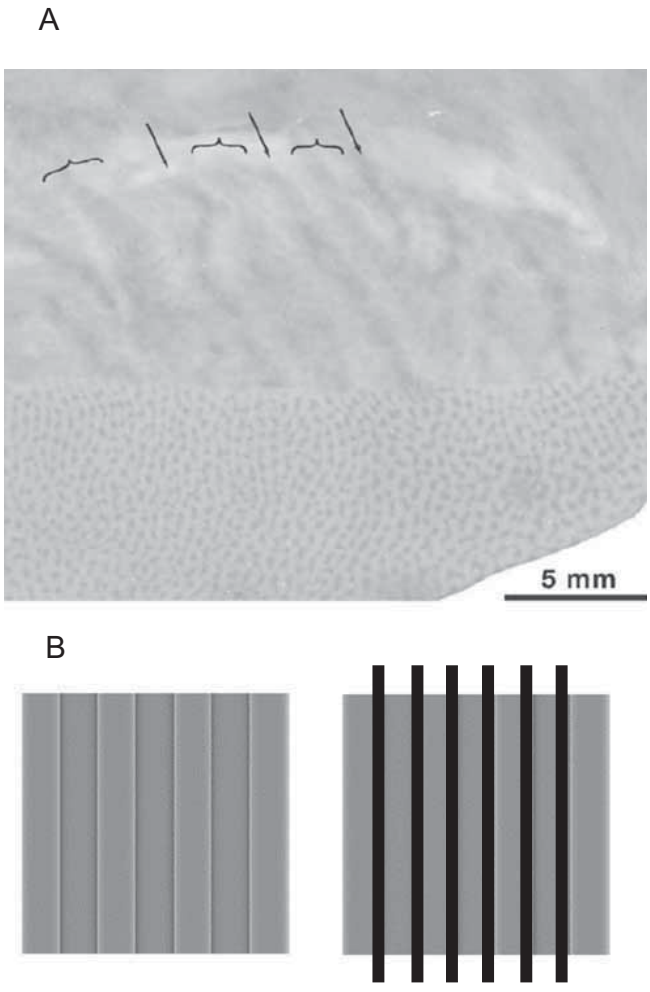


FIGURE 30.1 Cytochrome oxidase staining in V1 and V2, and Craik O'Brian Cornsweet stimulus. (*A*) Results from cytochrome oxidase staining in V1 (bottom half) and V2 (top half) reveal different anatomical subcompartments. Darker regions in V1 are blobs, lighter regions in V1 are interblobs. In V2, a striped pattern of staining can be discerned, consisting of thin stripes (arrows), thick stripes (brackets), and interstripes (least densely stained). (Reproduced from Sincich & Horton, 2005.) (*B*) Brightness illusion in the Craik O'Brian Cornsweet stimulus. The illusory brightness difference in the stimulus (on the left) is determined by the luminance at the edges of each patch. This can be appreciated by blocking the edges from view (on the right). Thin stripes (*A*) may play an important role in the perception of brightness spread as induced in the Craik O'Brian Cornsweet stimulus (see text).

clusters that are involved in spreading of surface properties are mixed with others that are involved in other functions. This has important implications for human functional magnetic resonance imaging (fMRI) studies, in which signal from neighboring clusters is pooled. This can make it difficult to link fMRI activity with specific neural computations during surface perception.

Recurrent loops in a hierarchical system Within each stream, there is evidence for *hierarchical processing*. A series of seminal papers (von der Heydt, Peterhans, & Baumgartner, 1984; Peterhans & von der Heydt, 1989; von der Heydt & Peterhans, 1989; Heider, Meskanaite, & Peterhans, 2000) indicated that V2 interstripe neurons integrate local measurements into neural signals that are relevant for contour perception, while such integration is absent in V1. Similarly, Roe and colleagues (2005) demonstrated that the activity of V2 interstripe neurons reflects the brightness percept in the Craik O'Brian Cornsweet (COC) stimulus (figure 30.1*B*), while V1 blob neurons do not. Hung, Ramsden, Chen, and Roe (2001) reported related data for the cat. The presence of activity in V2 more closely linked to perception than in V1 does not imply that processing takes place in a purely feedforward manner. Instead, *recurrent feedforward/feedback loops* are thought to validate local analysis in low-level areas by a more global analysis of the visual scene in higher-level areas.

A strong candidate for sending feedback relevant for surface perception to areas V1 and V2 is area V4, which maintains a segregation of processing streams emanating from V1 and V2 (Nakamura et al., 1993; Yoshioka & Dow, 1996; Xiao et al., 1999). V4 neurons have RFs with large surrounds (Desimone & Schein, 1987; Schein & Desimone, 1990), and they display complex stimulus selectivity, which suggests a role both in global aspects of boundary construction and figure-ground segregation (Pasupathy & Connor, 2002) and in the encoding of more complex aspects of surfaces, such as statistics of low-level features (Hanazawa & Komatsu, 2001; Tanabe, Doi, Umeda, & Fujita, 2005) that can help to determine perceived depth, slant, and curvature. V4 is therefore well placed to be involved in feedback that is relevant for both contour and surface networks in V1 and V2. First, with respect to contour processing, a recurrent loop involving V1, V2, and V4 may help to construct boundaries from local discontinuities and may assign special status to boundaries belonging to the foreground (border ownership), based on occlusion cues, stereo cues, and others (Zhou, Friedman, & von der Heydt, 2000; Qiu & von der Heydt, 2005). Second, with respect to surface processing, feedback from V4 that takes boundary status into account may determine whether a spreading process within boundary representations produces a "visible" surface or not (see the third part of the chapter). Although the processing levels in the model proposed in the second part of the chapter are limited to areas V1, V2, and V4, areas other than V4 are likely to provide relevant feedback to V1 and V2 (e.g., see Hupe, James, Girard, & Bullier, 2001; Hupé et al., 2001).

In summary, the strongest evidence for surface-related spreading processes points to thin-stripe modules in V2. Although a contribution of V1 to surface representation is likely, we hypothesize that spreading activation directly relevant for surface perception takes place predominantly in V2 thin stripes and related modules in higher-order extrastriate cortex (V4). The spreading is hypothesized to be

contained by inhibitory processes related to boundary representations in interstripe modules in V2 and related modules in V4. Feedback from area V4 to V2 and V1 is hypothesized to be instrumental for boundary representation and segregation as well as for the gating of spreading activity that leads to "visible" surfaces.

Evidence for Active Interpolation Theory in Early Visual Areas

Explicit activity related to surfaces Huang and Paradiso (2008) observed that monkey V1 neurons respond to large homogenous luminance surfaces overlapping with their RFs. Other studies combining single-unit recording and optical imaging demonstrated surface-related activity for COC stimuli in monkey V1 and V2 (Roe et al., 2005) and in cat areas 17 and 18 (Hung et al., 2001). These studies indicated that perceptual attributes of luminance-defined surfaces were encoded in extrastriate cortex. Interestingly, Perna, Tosetti, Montanaro, and Morrone (2005), using fMRI in human subjects, failed to find any correlate of perceived brightness in the COC illusion in lower-order visual areas. Mendola, Conner, Sharma, Bahekar, and Lemieux (2006) found nonretinotopic deactivation during static brightness filling-in in V1.

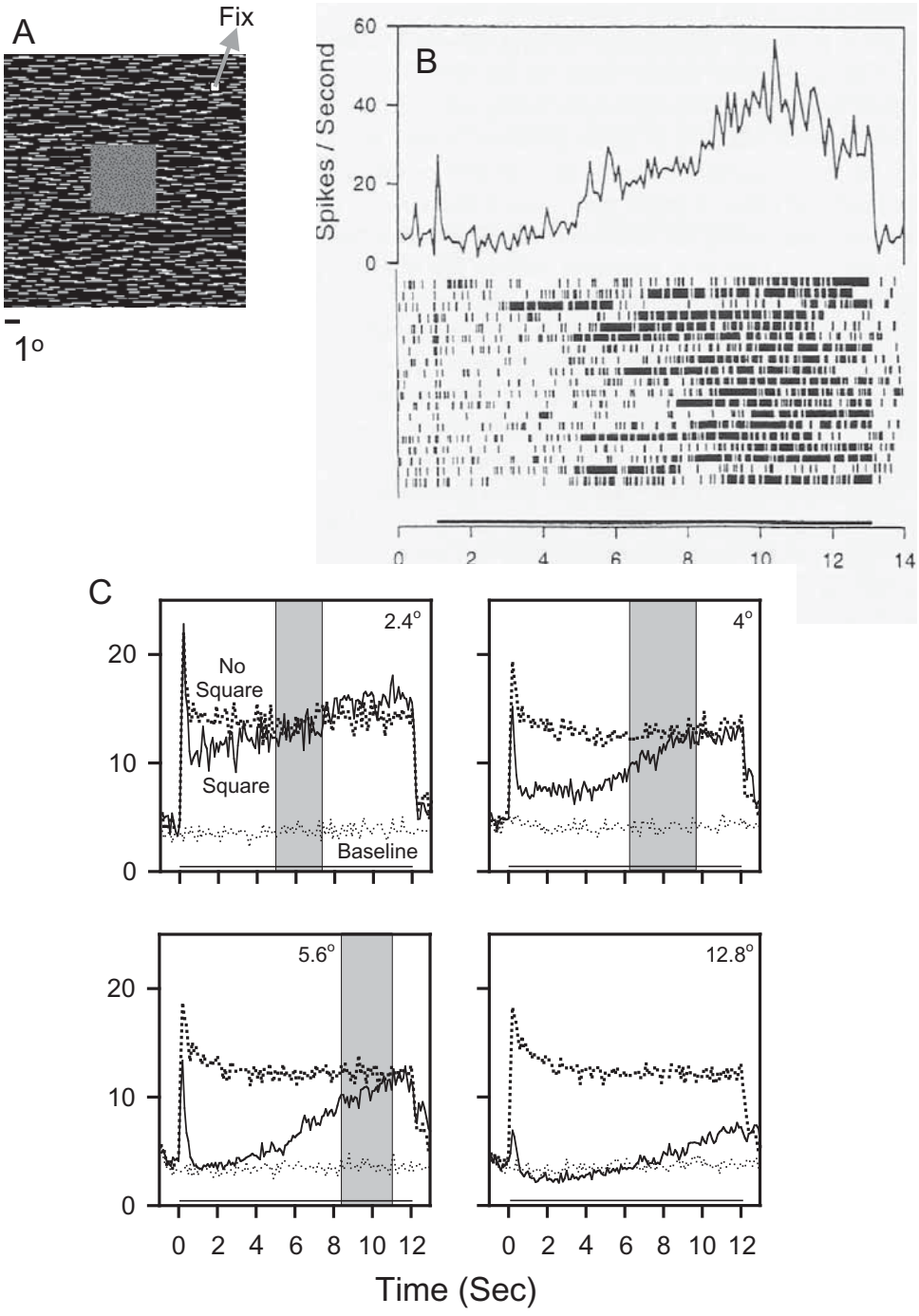
Several neurophysiological recording studies found neural correlates of brightness induction in single neurons of monkey V1 (Kinoshita & Komatsu, 2001) or cat area 17 (Rossi & Paradiso, 1996, 1999), although a recent fMRI study failed to replicate these findings in human subjects (Cornelissen, Wade, Vladusich, Dougherty, & Wandell, 2006; see the third part of the chapter). Neural responses related to color filling-in across the blind spot have been demonstrated in monkey V1 (Murakami et al., 1997; Matsumoto & Komatsu, 2005; Komatsu, Kinoshita, & Murakami, 2000), and using fMRI, Sasaki and Watanabe (2004) found a correlate of neon color spreading in human V1.

Using texture stimuli in a Troxler fading paradigm (figure 30.2*A*), De Weerd, Gattass, Desimone, and Ungerleider (1995) demonstrated increased responses in monkey V2 and V3 neurons whose RFs overlapped with a gray square surrounded by texture after several seconds of maintained fixation away from the square (figure 30.2*B*). When the response increase in monkey extrastriate neurons reached a level that was indistinguishable from activity obtained when the RFs were filled physically with the background, human subjects perceived filling-in under the same stimulus conditions (figure 30.2*C*). Furthermore, the time delay before filling-in of a gray figure by surrounding texture during maintained fixation away from the figure depended on the projection area in retinotopically organized visual areas (De Weerd, Desimone, & Ungerleiter, 1998) and was related to adaptation of an inhibitory signal (Tremere et al., 2003; De Weerd, 2006), possibly related to boundary adaptation. In line with findings with the COC stimulus (Roe et al., 2005), there were no activity increases in V1 neurons with RFs over the gray square, suggesting that V1 does not represent the perceptual quality of texture surfaces. Weil, Watkins, and Rees (2008) reported a reduction in fMRI activity in the V1 and V2 representations of a small, flickering target, while subjects reported filling-in of the target by surrounding dynamic texture.

Direct measurements of neural spread related to surface perception by recording from multiple electrodes in a surface's cortical representation have not yet been performed. However, several psychophysical studies managed to visualize spread of brightness (Paradiso & Nakayama, 1991; Paradiso & Hahn, 1996) and texture (Motoyoshi, 1999). Estimates of the speed of brightness spread based on psychophysical data have been on the order of 50 ms/degree in the COC stimulus (Davey, Maddess, & Srinivasan, 1998), and 10 ms/degree in the brightness-masking paradigm (Paradiso & Nakayama, 1991).

The data show strong neurophysiological support for neuronal correlates of brightness, color, and texture spreading in extrastriate areas. At the same time, support from fMRI studies is weak. One of the goals of our modeling effort is to achieve a better understanding of the divergences between fMRI and neurophysiological techniques.

Importance of boundary representations in controlling spread The role of boundary representations in controlling surface feature spread finds support in a number of studies. The optical recording data from Roe and colleagues (2005) in V2 and a similar study in the cat (Hung et al., 2001) confirm the importance of surface information at boundaries in determining surface percepts. Furthermore, a psychophysical study from Salmela and Laurinen (2007) suggests that brightness spread is facilitated when edges are sharp and that the blurring of edges prevents brightness spread. Hence full boundary representations including sufficiently high spatial frequencies (Marr, 1982; Morrone & Burr, 1988) might be required to induce and contain brightness spread within a figure region, and blurred boundaries might not permit adequate surface representation. Other studies using figures on a texture background in a Troxler paradigm confirmed the importance of local differences at the border (De Weerd et al., 1998; Stürzel & Spillmann, 2001; Yokota & Yokota, 2004, 2005; Welchman & Harris, 2001); larger differences tending to prolong the "survival" of the figure (and thus delaying the initiation of perceptual filling-in). Furthermore, blurring facilitates perceptual filling-in of a figure with its background (Zhang & von der Heydt, 1995; Friedman, Zhu, & von der Heydt, 1999; von der Heydt, Friedman, & Zhou, 2003), except when stimuli are placed too peripherally for blurring to have an effect (Welchman & Harris, 2001, 2003).



Direct evidence for the primacy of boundary representations comes from recording study in the cat (Hung, Ramsden, & Roe, 2007), in which responses from simultaneously recorded cell pairs were compared, one RF being over a surface boundary and another inside the surface. They observed a border-to-surface shift in the relative timing of spiking activity for both real and illusory (COC) brightness contrast stimuli. Interestingly, the difference between boundary and surface-related signals was observed predominantly in area 17–18 pairs and was weaker in area 17–17 pairs. Further, the reduction of inhibition observed in neurons with RFs in a figure surrounded by dynamic texture during Troxler fading (De Weerd, 2006) suggests that boundary adaptation permits the recorded neurons to become driven by excitatory input from the texture background. This is in agreement with the role of boundary representations in controlling neural spreading activation related to surface perception.

FIGURE 30.2 Stimulus used to induce Troxler fading and associated neuronal responses. (*A*) Dynamic texture stimulus with gray figure used by De Weerd and colleagues (1995, 1998) to induce perceptual filling-in of the figure by the textured surround. The picture shows a single frame of the dynamic texture stimulus. The homogeneous region in the center was approximately equiluminous with the average luminance of the surrounding dynamic texture and with the gray background upon which the texture was presented (23 cd/m^2). Since the position of the line elements was randomized in each frame of the movie, playing the movie (at 20 Hz) created a stimulus with continuously jittering line elements on the dark texture background. The small white square corresponds to the position of the fixation point (Fix). The illustration is approximately to scale except for the fixation spot, which is exaggerated in size for purposes of illustration. (*B*) Response from a single monkey V3 neuron with RF centered over the peripheral gray square (4 degrees) surrounded by texture, during passive fixation of a fixation spot away from the square (from Spillmann & De Weerd, 2003). The stimulus and square size shown in panel *A* was used, in which humans start to perceive filling-in approximately 6–10 s after stimulus onset. A recording trial started with the collection of 1 s of baseline activity in absence of a stimulus. Stimulus presentation is indicated by the solid bar at the bottom of the graph. The cell's RF straddled the square's edges, causing a brief onset response. The top panel shows the cumulative histogram demonstrating increased activity toward the end of stimulus presentation. Activity is expressed as spikes per second, calculated in bins of 100 ms. Individual response traces for the individual trials are shown below the cumulative histogram. Each small vertical mark indicates an action potential (at high firing rates the vertical marks tend to blend). Individual trials show discrete episodes of increased activity, which were more likely to occur toward the end of each trial. The histogram only includes trials in which the monkey maintained fixation till the end of the trial. (*C*) Correlation between increased neural activity in V2/V3 and perceptual filling-in (from De Weerd et al.,1995). Average activity in stimulus conditions with a gray square over or inside the RF (solid curve) and in conditions in which the square, and thus the RF were physically filled with texture (heavy dotted curve) in subsets of V2 and V3 neurons with significant activity increases. Fine dots show baseline activity recorded without a stimulus. Square size is shown at the top of the panels. The horizontal line on top of each abscissa indicates the 12-s stimulus presentation time, preceded and followed by a 1-s period in which activity was recorded in the absence of a stimulus. Human observers reported filling-in for the same stimuli as were used during physiological recordings in the monkey in a time range indicated by the shaded zones. Activity increases were statistically evaluated in 93 V2/V3 neurons that were recorded in all square size conditions, using paired *t*-tests (two-tailed, $p < 0.05$), in which firing rates in a 2.5-s interval starting 1.5 s after stimulus onset were compared with the rates in the last 2.5 s of stimulus presentation. For cells that initially responded similarly in the square and no-square conditions, there was less opportunity to show an activity increase. This was particularly true at square sizes of 3.2 degrees and smaller (which were also not as conspicuous to human subjects as the larger square), for which we found significant activity increases in only 12% of V2/V3 cells. For larger square sizes, significant activity increases were found for about one-third of V2 neurons and about half of V3 neurons. Thus different numbers of neurons and trials (up to 20 trials per condition per neuron) were included in the different histograms at different square sizes. The inhibition during presentation of the largest square in the early part of the response (1–4 s after stimulus onset) was predictive for the magnitude of response increases in the late part of the response for smaller square sizes. This suggests that adaptation of inhibition may be a permissive factor for the observed response increases in the recordings with RF over the square (De Weerd, 2006).

Surface spreading processes are controlled not only by boundary representations, but also by more global aspects of the stimulus. Data from Troxler paradigms suggest that factors that determine figure-ground assignment (Sakaguchi, 2001, 2006; De Weerd et al., 1998; Hamburger, Prior, Sarris, & Spillmann, 2006; Hsieh & Tse, 2006) and statistics of the textures themselves (Hindi Attar, Hamburger, Rosenholtz, Götzl, & Spillmann, 2007; Sagakuchi, 2006) play an important role in determining the outcome of the spread of visual surface features.

Overall, the data suggest that boundary representations control spreading activation, both by initiating and containing spread in normal vision and by permitting new spread when boundaries adapt during stabilized vision.

EMPIRICAL BASIS OF A COMPUTATIONAL MODEL OF SURFACE FILLING-IN According to our interpretation of the data, four principles of visual system organization can be used to guide the modeling of perceptual filling-in. First, surface representation is accomplished by separate but interacting boundary and surface mechanisms that are intertwined in V1, V2, and V4. Second, spread of surface properties occurs via lateral connectivity in the surface representation system. Third, recurrent loops within hierarchically organized border and surface processing streams determine the perceptual outcome of spreading activation. Fourth, spread of surface feature is contained by inhibitory signal from the boundary system, and adaptation of inhibition is a permissive factor for new spreading activation across surface boundaries.

Rather than focusing on luminance and color stimuli (e.g., Grossberg, 2003a; Neumann, Pessoa, & Hanson, 2001), the present model aims to simulate texture filling-in in a Troxler paradigm, in which a figure is "invaded" by the background following figure boundary adaptation. We will not consider initial spreading events at stimulus onset. We aim to increase understanding of divergent results from fMRI and spiking data in filling-in studies, to study implications of anatomical intertwining of boundary and surface processing streams for the fMRI activity distribution, and to increase insight into the contribution of recurrent loops to surface perception.

Structural Assumptions and Architecture of the Neural Network Model

Properties of elementary, single processing units A single processing unit in the neural network is not related to a single neuron but represents a whole cortical column (Goebel, 1993). Cortical columns can be considered the building blocks of neural coding within brain areas (e.g., Hubel & Wiesel, 1959, 1962). A single processing element in our network is called a *cortical column unit* (CCU). The dynamics of a single CCU is described by the following standard difference equations:

$$net_i(t) = \sum_{j=1}^{n} w_{ij}^{a} a_j$$

$$a_i(t) = (1-r)\,a_i(t-1) + \tau\sigma\left(net_i(t) + b_i\right)$$

where w_{ij} is the weight from unit j to unit i, $a_i(t)$ is the average spike output of unit i at time t, $net_i(t)$ is the net input (excitatory minus inhibitory input) for unit i at time t, b_i is a bias term, and $\sigma(x)$ is the logistic (sigmoidal) function. The value τ ($0 < \tau \leq 1$) determines how strongly the activation value (average spiking activity) at the last time point ($t-1$) is influencing the activity at the current time point t.

Global architecture of the neural network model Our large-scale recurrent neural network model (figure 30.3) incorporates *retinotopically organized*, rectangular sheets representing areas V1, V2, and V4. To distinguish simplified *Model* regions from related visual areas, the modeled regions are denoted as *Model*-<area> or M-<area>. At present, the network processes only luminance-defined input, and no distinction is made between parvocellular and magnocellular processing. The luminance input to the model retina (M-Retina) is encoded as a spatiotemporal pattern across a single rectangular input sheet. Although we demonstrate processing with a network of 80 × 80 M-Retina units, the network can be scaled programmatically to any desired size. Details of the model architecture and its operation are provided at http://www.brainvoyager.com/n3d/finn/index.html.

A CCU in M-V1 integrates activity from a small retinal patch forming classical (afferent) RFs. The sampling size of retinal patches increases from central to peripheral regions so that the overall connection pattern from M-Retina to M-V1 reproduces cortical magnification (e.g., Sereno et al., 1995). By further convergence, units in M-V2 and M-V4 *hierarchically* integrate responses from lower levels, creating larger RFs (figure 30.3). *Lateral connections* within M-V1, M-V2, and M-V4 as well as *top-down connections* from higher visual areas (M-V4 → M-V2; M-V2 → M-V1) implement contextual effects from outside the classical RF.

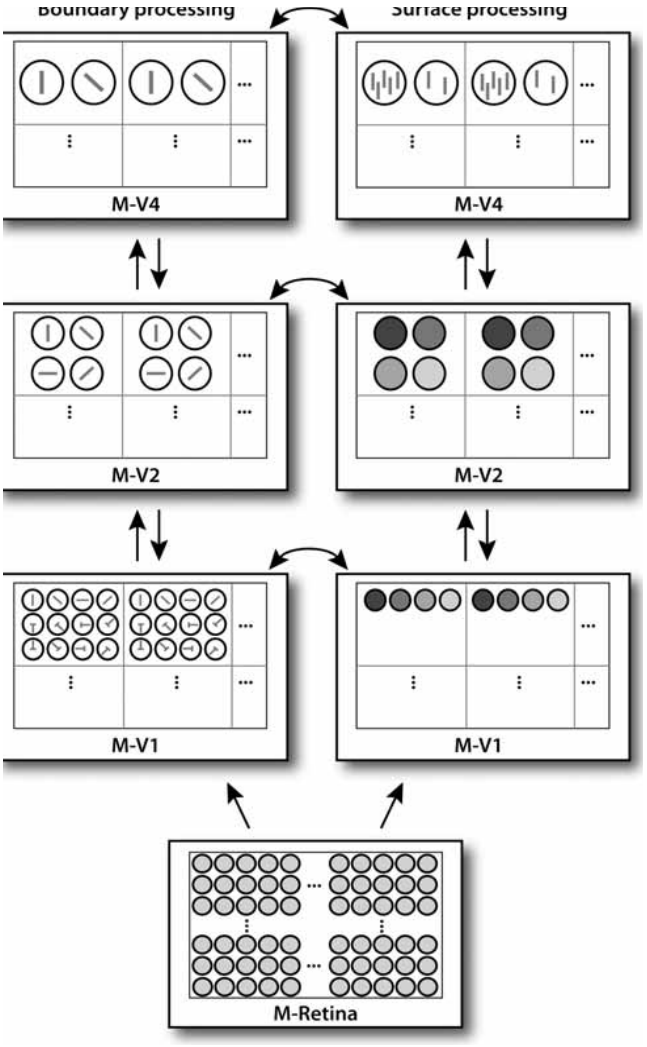


Figure 30.3 Architecture of the model. The model retina (M-Retina, 80 × 80 units) is the source of two streams specialized for boundary processing and surface processing, respectively, extending along three modeled areas, M-V1 (interblobs, blobs), M-V2 (interstripes, thin stripes), and M-V4. The connection pattern from M-Retina to M-V1 implements cortical magnification. Receptive field sizes of the retinotopically organized areas M-V1, M-V2, and M-V4 increases with increasing distance from M-Retina as indicated by the increasing size of the depicted units (circles). At each position, a number of different feature cells (hypercolumn) analyze a small, topologically corresponding, region at a lower level. As indicated by icons within the depicted units, the boundary-processing system detects oriented lines and end-stops, while the surface processing system contains units for luminance/brightness detection (M-V1/M-V2) and texture units (M-V4). Lateral and recurrent connections produce dynamic interactions within and between the boundary- and surface-processing systems. These connections include inhibitory synapses from each unit in the boundary-processing system to topologically corresponding units in the surface-processing system to prevent suprathreshold lateral spread in the surface system.

An important feature of the model is the definition of two parallel, interacting streams, *a boundary-processing system* and *a surface-processing system*, within each of the three relevant hierarchical levels. The boundary-processing subparts of M-V1 (interblob), M-V2 (interstripe), and M-V4 contain four orientation units at each position. The boundary processing subparts of M-V1 (interblobs), M-V2 (interstripes), and M-V4 also contain at each position 4×2 units with RFs detecting oriented line endings in two directions (bipolar "end-stop" units). For better visibility, units that code the same feature are shown as two-dimensional layers, separated from layers coding other features (see figures 30.3 and 30.6). A single feature map contains 30×30 units (see figure 30.6). For an anatomically more correct arrangement and to visualize modeled aggregate fMRI activity, different layers can also be stacked behind each other. A single set of feature detectors for the same small retinal position is referred to as a *hypercolumn*. The surface-processing subparts of M-V1 (blobs), M-V2 (thin stripes), and M-V4 contain units for the representation of surface features. In the case of homogeneous surface material, a unit in the surface-processing system represents luminance-related activity pooled from a small area of M-Retina. In the case of textured surfaces (e.g., a large field of small, randomly placed vertical lines), a M-V4 texture unit responds to a small surface patch if it contains a few elements at a preferred orientation. V4 contains texture units for a continuous range of orientations and texture element densities, but here we have included only layers in the model to demonstrate texture filling-in for vertically aligned texture elements. The model permits lateral spreading of different features at the same time in different M-areas (e.g., Ramachandran & Gregory, 1991). Further, *recurrent connections* produce dynamic interactions within and between the boundary-processing and surface-processing systems that determine the perceptual outcome of surface-related spreading processes.

The interaction between boundary and surface systems is modeled by the inhibitory connection of units in the contour processing system onto topologically corresponding units in the surface processing system. Boundary adaptation in the Troxler paradigm is modeled by a negative self-recurrent connection, causing activity in contour units to gradually decay over time. This decay reduces the inhibitory effect on the surface processing system eventually permitting *active* lateral spreading. These interactions occur simultaneously at multiple levels.

Network-brain links to visualize spiking and fMRI activity BOLD data reflect (suprathreshold) spiking activity, hemodynamic spread, and neural spread of subthreshold neural activity (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Logothetis & Wandell, 2004; Oelterman, Augath, & Logothetis, 2007; Maier et al., 2008). To investigate discrepancies between spiking and BOLD data, spiking and the other components of BOLD must be modeled separately in an environment that permits a comparison with matching empirical data (Goebel & Horwitz, 2008). To compare modeled activity distributions with empirical data, modeled hypercolumns can be linked to topographically matching voxels obtained from structural brain scans and retinotopic mapping in human subjects, thereby establishing a common representational space for simulated and measured data. Thanks to these network-to-brain (NB) links, predicted data (e.g., fMRI signal) can be analyzed by using the same analysis tools as are used for the measured data (e.g., General Linear Model, Multi-Voxel Pattern Analysis). As an illustration, figure 30.4 shows NB links as established between M-V1 and topographically corresponding visual cortex of a subject.

To enable a direct comparison of modeled synaptic activity and spiking activity, the activity state of a unit is visualized in three ways using either the integrated synaptic activity net_i^{abs}, the average spike output level a_i, or the simulated fMRI data y_i. Note that net_i^{abs} is different from the standard "netto" input *net*, since it calculates the sum of the *absolute* activity arriving via excitatory and inhibitory synapses at unit *i*: $net_i^{abs} = |\Sigma \omega_{ij}^{ex} a_j| + |\Sigma \omega_{ij}^{inh} a_j|$. To create simulated proto-fMRI data in neuronal timing, the integrated synaptic input signal of a unit and the average spike output level is integrated as

$$a_i^{fMRI} = \lambda net_i^{abs} + (1-\lambda) a_i$$

The weighting constant is set to a value of $\lambda = 0.8$, thereby biasing the fMRI signal to reflect the synaptic rather than the spiking activity. The ultimately predicted fMRI output value y_i is obtained by convolving a hemodynamic response function HRF(x) with the predicted proto-fMRI data a_i^{fMRI}:

$$y_i = \mathrm{HRF}\left(a_i^{fMRI}\right)$$

For the HRF function, a single-gamma function (Boynton, Engel, Glover, & Heeger, 1996) or a two-gamma function (Friston et al., 1998) may be chosen in our model environment.

Simulation Studies

Mapping a gray figure in a dynamic noise texture background Prior to any fMRI investigation of perceptual filling-in of a figure by its surrounding background in a Troxler paradigm, it is essential to demonstrate that fMRI has the spatial resolution to distinguish the two surfaces. As is shown in figure 30.5, De Weerd, Karni, Kastner, Ungerleider, and Jezzard (1997) reported a failure in distinguishing fMRI activity obtained in V1 with a full-field dynamic texture from activity obtained with a texture containing a square figure, with figures as large as 6×6 at 7.2

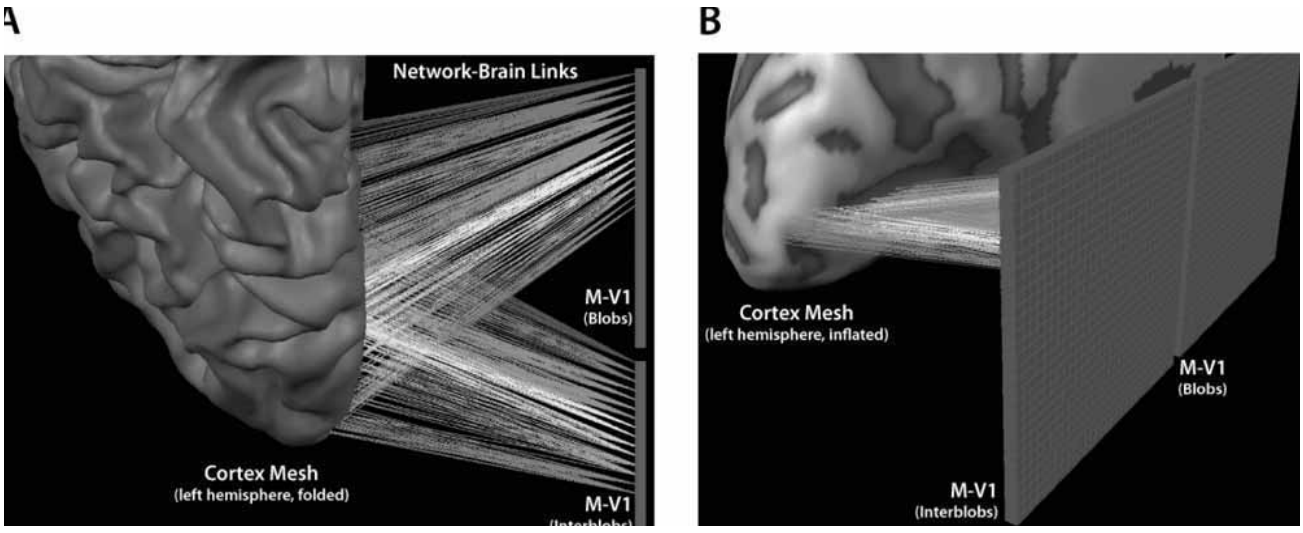


FIGURE 30.4 Selected network-brain (NB) links connecting cortical column units from one layer of the boundary processing system (interblobs) and one layer of the surface processing system (blobs) of M-V1 to a left hemisphere cortex mesh of a subject. Colored lines indicate NB links from foveal (red) to increasingly peripheral (orange, yellow, green, blue) locations within the model layers and corresponding regions within the calcarine sulcus of the subject's cortex as identified by a fMRI retinotopic mapping experiment. (*A*) Top view depicting the cortex mesh in a folded state. (*B*) Lateral view depicting the cortex mesh in an inflated view. Note that NB links originating from the same retinotopic position of the boundary-processing M-V1 layer and surface-processing M-V1 layer are connected to the same position at the cortex mesh. (See color plate 43.)

degrees of eccentricity (corresponding to about 10 mm in V1). This was surprising, because estimates of spatial resolution on the order of 2 mm had been reported around the same time (e.g., Engel et al., 1994; Engel, Glover, & Wandell, 1997). On the basis of our own findings, we have suggested that spatial resolution was much lower (Gaussian point spread of 7 mm at HWHM; De Weerd et al., 1997). The data from De Weerd and colleagues (1997) were obtained by using a 1.5T scanner and surface coil, but in recent years, we have confirmed their observations using a 3T scanner (unpublished data).

To investigate this phenomenon, we conducted a simple simulation study in which we presented either a rectangular figure made of dynamic texture on an "empty" background (Fig-On-Back-Off) or the inverse stimulus with an empty rectangular figure on a dynamic texture background (Fig-Off-Back-On). The figure was varied in size, and the resulting activity levels in the background and figure are shown in figure 30.6 (active voxels in white). Modeled fMRI signal is shown in M-V2, but it is reasonable to expect similar limitations in spatial resolution in other extrastriate areas. Modeled fMRI activity for a small empty figure on a texture background was as high in the figure as in the background representation (figure 30.6*A*). Only for a large empty square, the fMRI signal revealed a gradual fall-off of activity from the texture toward the center of the empty figure (figure 30.6*B*). Modeled spiking activity, however, was elevated only in the background representation and did not invade the representation of the empty figure (figures 30.6*C* and 30.6*D*). The simulated, *passive spread of fMRI* may reveal the spread of *subthreshold synaptic activity* into modeled cortical regions where spiking activity is absent (for discussion, see the third part of the chapter). As expected, there was also significant spread of modeled fMRI signal from a stimulated figure representation into an empty background representation (Fig-On-Back-Off; figures 30.6*E* and 30.6*F*) beyond the activity distribution defined by spiking activity (figures 30.6*G* and 30.6*H*).

Perceptual filling-in of a gray figure by a dynamic noise texture background The lateral spreading described in the previous section does not lead to spiking activity in surface-processing units within the representation of a homogenous figure surrounded by texture. Hence the modeled fMRI spread is not a correlate of perceptual filling-in but rather a phenomenon that presents an obstacle to measure a fMRI correlate of perceptual filling-in. Here, we model filling-in of a figure by surrounding fine-grained dynamic noise. Prior to filling-in, inhibitory influences from the boundary-processing system are thought to contain spiking activity within surface boundaries. Figure 30.7*A* shows *spiking* activity during this state in boundary and surface systems in M-V1 and M-V2 (activity in white), during the presentation of a homogenous figure on a dynamic noise background to the M-Retina. Figure 30.7*B* shows fMRI activity for the same system state in boundary and surface systems of M-V2. The figure is sufficiently large to prevent the inflow of passive fMRI activity from the background to reach the middle of the figure rep-

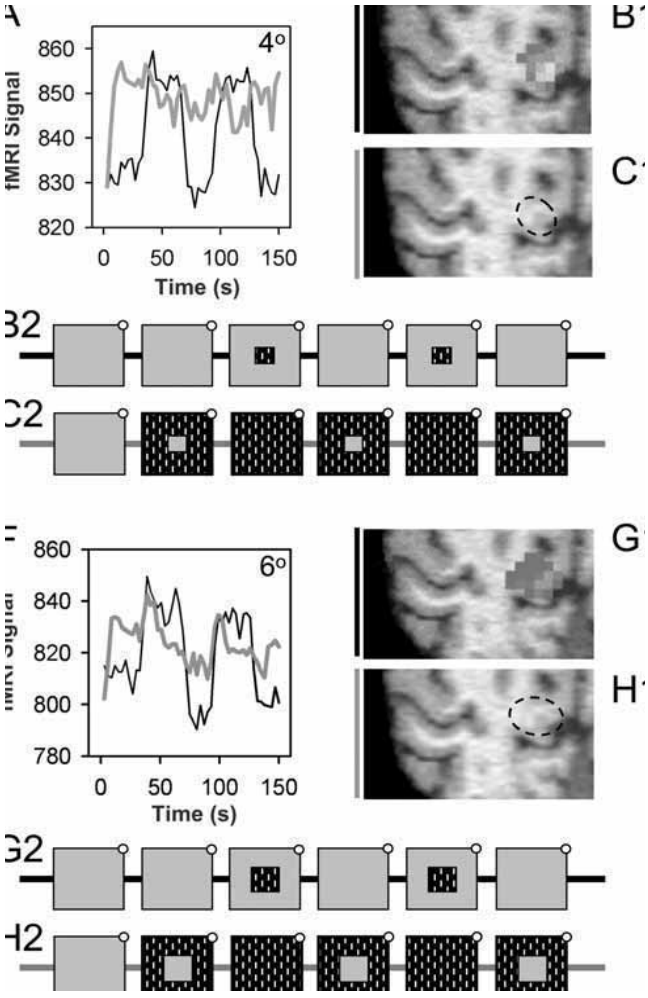


FIGURE 30.5 Empirical data from a single subject illustrating limitations in fMRI resolution using 1.5 Tesla Signa Horizon Echospeed system. Similar data were obtained from a second subject, and data in both subjects were replicable across sessions. Functional scans were obtained with a gradient echo EPI sequence (BOLD images), using a 64 × 64 matrix, a FOV of 14–16 cm, coronal slices with Thk = 4 mm, TE = 40 ms, and TR = 4 s. Functional data were overlaid on high-resolution structural scans of the same person's brain. Structural scans were obtained by using 3D-SPGR, a 512 × 384 × 128 matrix, with TE = 5 ms, TR = 24 ms, and a flip angle of 45 degrees. The testing of the effects of interest resulted in Wilkinson's maps, which were converted into *z*-maps. Single voxels were considered significant when the corresponding *z*-score exceeded 3.07. The coronal slice shown was positioned 16 mm anterior from the occipital pole. Dynamic texture stimuli (see figure 30.1*A*) were equiluminant with the background (24 cd/m^2). (*A–C*) fMRI signal as a function of time (*A*) during two block designs (*B*,*C*). In the first design (*B*2), two 30-s blocks of presentation of 4 degree square at 7.2 degrees eccentricity in lower left quadrant were interleaved with three 30-s periods of gray background. In the second design (*C*2), two 30-s blocks of presentation of a textured background were alternated with three 30-s periods of presentation of a gray 4-degree square at 7.2 degrees eccentricity. An initial period of baseline measurement without stimulus was discarded from analysis. Fixation spot (where subjects performed a demanding T/L discrimination task) is indicated at the top right of each stimulus panel. When the square was defined by dynamic texture on a gray background (*B*2), significant activity (black plot in *A*) was found in a large number of voxels in the upper bank of the calcarine sulcus (*B*1) using a regressor corresponding to the timing of the texture square in the block design in panel *B*2. When a full texture background was shown in blocks with or without a gray square (*C*2), there was no significant activity (*C*1) for the regressor corresponding to the physical filling-in of the gray square with texture (*C*2). The gray plot in panel *A* shows activity in the region of interest (ROI) defined by the response to the textured square (ROI is shown as a dashed oval in *C*1) during data from the block design shown in *C*2. For the design in *C*2, the data suggest spread of fMRI signal from the background texture into the gray figure. (*F–H*) Same conventions as in panels *A–C*, but the square size was 6 degrees. The gray plot in panel *F* shows a small, transient response to physical filling-in of the texture (design *H*2) that did not lead to significant activity in *H*1 (based on regressor corresponding to design *H*2).The fMRI signal in designs *C*2 and *H*2 inside the figure representation was not due to perceptual filling-in, as the signal was present from the beginning of stimulation (gray plot in panels *A* and *F*). Based on data in block designs showing a texture square on a gray background alternated with a gray background (as in *B*2 and *G*2), we found (averaged over subjects) activated regions of 158, 230, and 267 mm^2 for square sizes of 1, 4, and 6 degrees, respectively, while based on retinotopy (Sereno et al., 1995), activated regions of approximately 3, 45, and 110 mm^2 were expected. Averaged over subjects and conditions, a Gaussian filter of 7 mm (HWHM) was required to simulate the blurring of expected signal by fMRI. (See color plate 44.)

resentation and thus allows for the possibility to measure an fMRI signal related to perceptual filling-in exceeding fMRI signal related to passive spread. Figure 30.7*C* shows fMRI activity associated with adaptation of boundary units and ensuing perceptual filling-in of the figure representation by activity from the background. Adaptation of the boundary units is implemented by self-recurring inhibitory connections. This causes an increase in net excitation of units within surface modules in the figure representation closest to the boundary. After some time, the units reach their threshold, producing spiking activity. Because of the excitatory lateral connections within the surface-processing system, units farther inward in the figure representation start to receive additional excitatory input from their spiking neighbors, which leads to spiking activity also in these units. Because of this evolving chain reaction, spiking activity quickly spreads inward until all units within the figure representation exhibit increased spiking activity. At this stage, the activity level of units within the former figure representation is comparable to the activity level that would have occurred had texture stimulus without a figure been presented. This activity state in the M2 surface-processing system is the modeled correlate of perceptual filling-in.

Perceptual filling-in of a homogenous figure by a dynamic, coarse texture background Many studies of perceptual filling-in use coarse

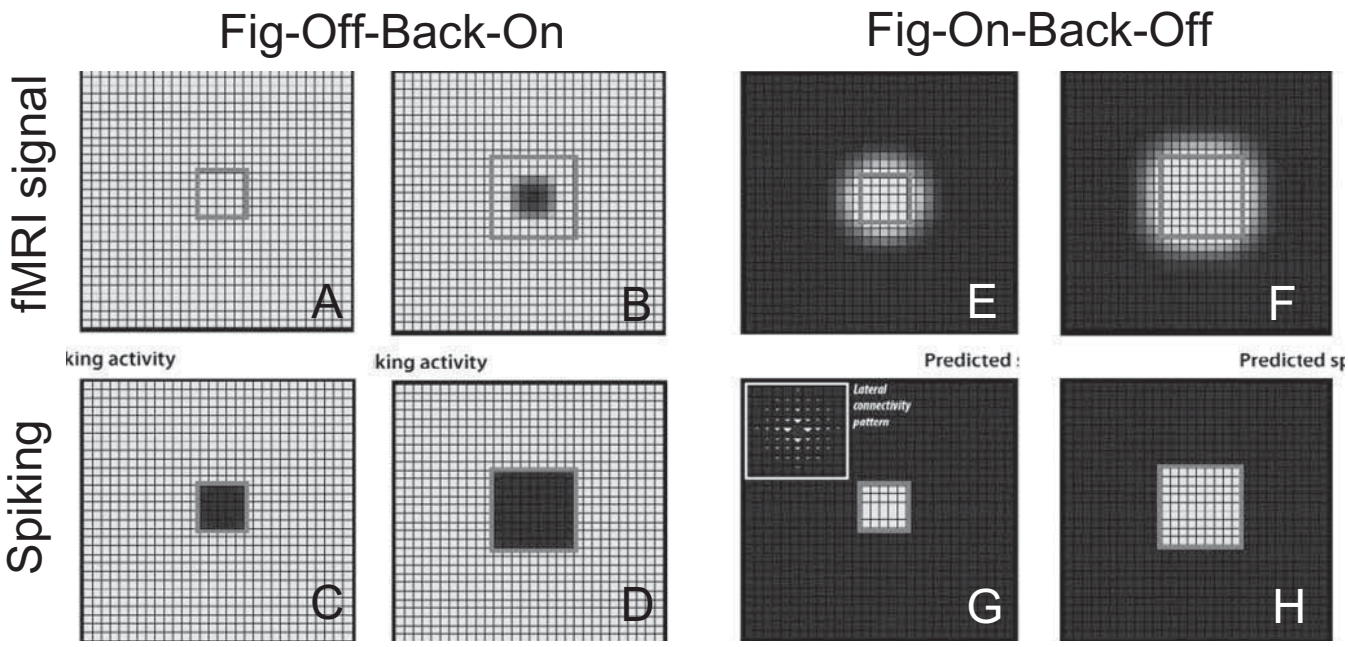


FIGURE 30.6 Simulation of spiking activity and fMRI signals when an "empty" square is surrounded by dynamic texture (Fig-Off-Back-On) or vice versa (Fig-On-Back-Off). As indicated by red outlines, stimuli were presented with a small square (5 × 5 rectangle; *A*, *C*, *E*, *G*) or a large square (9 × 9 rectangle; *B*, *D*, *F*, *H*). Each panel shows the activity state of the same layer from the surface-processing system of M-V2. The lateral connectivity pattern of each unit within this layer is shown in G. The activity state of a processing unit is indicated by a black-to-white color range corresponding to weak-to-strong activity. The predicted fMRI data in the Fig-Off-Back-On condition show inflow from texture background into the figure representation (*A*, *B*), and only for the larger square (*B*) the interior is spared (shown in dark). Predicted spiking data (*C*, *D*) show a perfect representation of the square, without noticeable inflow from the background. The predicted fMRI data in the Fig-On-Back-Off condition show outflow of fMRI activity from the representation of the texture square into the empty background (*E*, *F*), while such outflow is unnoticeable for spiking data (*G*, *H*). (See color plate 45.)

texture backgrounds (e.g., De Weerd et al., 1998; Hindi Attar et al., 2007). For textures, especially when they are coarse, it can be asked what information is actually spreading during filling-in and how it is computed. In the first part of the chapter, we reviewed evidence suggesting that V4 neurons encode global statistics of texture patches, such as overall brightness, brightness, gradients, and texture density (Hanazawa & Komatsu, 2001; Tanabe et al., 2005). Similar statistical operations on multiple elements within a RF have been reported for V2 neurons (Anzai, Peng, & Van Essen, 2007), but unless textures consist of fine elements that are very densely packed, RFs of neurons in V2 might be too small to produce reliable estimates of texture statistics. Cortical areas at a level higher than V4 probably would produce statistical estimates that are insufficiently local to guide spreading processes in lower-order, retinotopic visual areas. Hence in our model, we assume that M-V4 has a special role in estimating global texture statistics and that these statistics are the kind of information that spreads during perceptual filling-in.

The computational processes that lead to spreading activation in figure representations in retinotopic maps are the same as those described for fine-grained textures, but they now involve interactions between boundary- and surface-processing systems within both M-V4 and M-V2 (for details, see http://www.brainvoyager.com/n3d/finn/index.html). In addition, a recurrent loop involving the two M-areas is required to produce spreading activity in either area. Before active neuronal spreading has occurred, inhibition from units in the boundary-processing system of M-V4 is strong enough to prevent lateral spreading within the surface-processing system of M-V4, and the same holds in M-V2. After boundary adaptation, active spreading within M-V4 and M-V2 may occur, but the activity levels in units within the surface-processing systems of M-V4 and M-V2 are codependent. More specifically, lateral subthreshold inputs to units in the M-V4 surface module inside the representation of the homogenous figure must be supplemented with *feedforward* subthreshold input from units in M-V2 surface module at corresponding retinotopic locations. Similarly, lateral subthreshold inputs to units in the M-V2 surface module inside the representation of the homogenous figure must be supplemented by *feedback* subthreshold input from units in the M-V4 surface module at

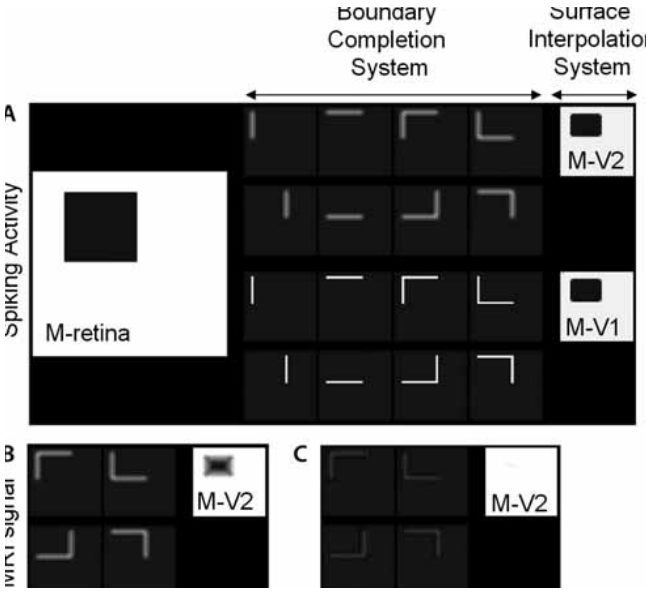


FIGURE 30.7 Filling-in simulation for an "empty" square surrounded by a fine noise texture. Average spiking or fMRI activity in relevant layers of the model is indicated on a dark-to-light scale (corresponding to low-to-high activity). (*A*) Initial spiking activity state after presenting the stimulus at M-Retina (left hand, large sheet). The boundary processing systems in M-V1 and M-V2 detect the borders separating the background from the central hole. The surface-processing system in M-V1 and M-V2 responds strongly to the texture background, revealing the empty square representation by an absence of activity. (*B*) fMRI activity in boundary and surface systems, with some spreading of fMRI signal into the figure representation in the surface system. (*C*) fMRI activity showing adaptation of boundary representations and ensuing active spreading in the surface system of M-V2. In panels *B* and *C*, only layers of M-V2 are shown, since the activity state in other layers do not change substantially.

corresponding retinotopic locations. This results in a recurrent loop, which leads to spreading of spiking activity from the texture surround into the figure within both M-V2 and M-V4.

By building appropriate structural assumptions into the recurrent M-V2/M-V4 loop, it could be possible to model effects of surface properties on perceptual filling-in during Troxler paradigms (e.g., see Hindi Attar et al., 2007). We focused on the contribution of recurrent loops to the interpolation of coarse texture, but such loops likely play a role in interpolation of other surface features. For particular features, such as brightness and color, recurrent loops that are relevant for surface interpolation might involve M-V1.

Insights, limitations, and future research directions

A SUMMARY OF ASSUMPTIONS AND INSIGHTS The main empirically based, structural assumptions that were built into the model's global architecture were (1) the presence of three levels of processing (M-V1, M-V2, and M-V4), (2) separate but interacting boundary- and surface-processing streams, (3) lateral connectivity within each module at each level, (4) recurrent connectivity between levels within modules, and (5) a specific strength of recurrent inhibition within the units of the boundary-processing system.

Below, we highlight three insights that follow from our modeling effort. First, we discuss consequences of the sensitivity of the fMRI signal to subthreshold neural activation for experiments on perceptual filling-in, for the notion of point spread, and for the comparability of fMRI and neuronal recording experiments. Second, we consider consequences of the inhomogeneity of cortex for the interpretation of fMRI data on perceptual filling-in. Third, we emphasize that the perceptual consequences of the recurrent loops in our model are in agreement with recent views on visual awareness and "conscious" perception.

INSIGHTS FROM MODELING THE SENSITIVITY OF FMRI TO SUBTHRESHOLD NEURAL ACTIVATION The explicit computation and visualization of average spiking activity and predicted fMRI data has proven valuable to understand the difficulty of mapping a homogenous figure surrounded by dynamic texture in retinotopic areas with fMRI. In the modeled data, we observed that a homogenous figure can easily be detected in the distribution of spiking activity while remaining undetectable in the distribution of fMRI activity. Our model thus reconciles the spiking results of De Weerd and colleagues (1995) with the fMRI data from De Weerd and colleagues (1997). Both studies investigated neural activity in the cortical representation of homogenous figures surrounded by texture. The former study found low spiking activity inside homogenous figures (prior to perceptual filling-in), whereas the latter study suggested that unless very large figures were used, the fMRI signal in the figure representation was high from stimulus onset on, irrespective of whether the figure was present or not. Our model might help to reconcile other, apparently conflicting results from fMRI studies and spike recording studies in animals that investigated the same phenomena across species (e.g., compare Kaas, Collins, & Chino, 2003, with Smirnakis et al., 2005; and see Tolias, Smirnakis, Augath, Trinath, & Logothetis, 2001; Maier et al., 2008). Note that while spiking data can be dissociated from fMRI signal, local field potentials (LFPs) are more robustly linked to the fMRI signal, as both LFPs and fMRI signal are thought to reflect predominantly (subthreshold) synaptic activity from large neuronal populations (Logothetis et al., 2001; Maier et al., 2008).

Our data also illustrate that estimates of fMRI spatial resolution, which have been reported to be as low as 1.8 mm

HWHM of Gaussian point spread (Engel et al., 1994, 1997), are often uninformative with respect to the prediction of fMRI activity distributions for stationary figures in a context of other stimuli. In fMRI studies that estimate spatial resolution, slowly moving dynamic checkerboard stimuli are used, presented isolated on a homogenous background. The isolated nature and the slow motion of these stimuli could limit the neural spread of subthreshold activity and reveal predominantly hemodynamic spread, thereby leading to the suggestion of excellent spatial resolution. However, complex stimuli composed of several parts that perceptually group together, or perceptually fill in, and involving prolonged stimulation at constant locations may increase the subthreshold inputs to neurons with RFs in between stimulus parts. This can lead to an *effective* fMRI spatial resolution (point-spread of spiking, hemodynamic, and subthreshold neural spread combined) that may be much lower than expected. Under these conditions, estimates have been reported on the order of 7 mm (De Weerd et al., 1997; Cornelissen et al., 2006). The sensitivity of fMRI signal to subthreshold neural activation makes fMRI point spread dependent on stimulus design and renders the notion of a generally applicable estimate of fMRI point spread obsolete.

Insights from Modeling Subvoxel Specialization of Visual Cortex The subvoxel specialization among and within compartments of early visual cortex, combined with physical limits on temporal and spatial resolution, renders an fMRI correlate of surface-related spreading activity difficult to detect. At the neuronal level, spreading activity related to surface encoding (e.g., in V2) is present only in small subsets of neurons that are intertwined with other neurons that do not participate in the spreading or might be involved in other functions. Moreover, boundary-related effects that might intuitively be considered local might spread into the representation of the figure because of the sensitivity of fMRI to passive spread of subthreshold activity. Hence decreased responses to boundaries could contribute to the overall activity measured from the figure representation. If this signal decrease is relatively strong, it might outweigh any activity increases associated with surface-related spreading, and one might conclude incorrectly that surface interpolation is associated with a decrease in activation. Thus even in studying filling-in of large stimuli, the effective spatial resolution of fMRI might be insufficient to detect a weak correlate of filling-in and might instead reveal the stronger boundary adaptation effect that passively spreads into the figure representation (Mendola et al., 2006, Perna et al., 2005). The study from Weil and colleagues (2008) in which filling-in of a flickering target was associated with a decreased fMRI signal also suffered from these limitations: The 1.2 × 1.2 degree target at 8.75 degrees of eccentricity used in that study corresponds approximately to 1.3 × 1.3 mm of V1 cortex (Sereno et al., 1995), well below the 3 × 3 mm voxel size that is used for functional scanning. It cannot be excluded that reduced activation associated with adaptation to the (boundaries of) the flickering target dominated increased signal related to perceptual filling-in. In the work of Cornelissen and colleagues (2006), the failure to confirm a correlate of brightness induction might have been due to a combination of insufficient spatial and temporal resolution to resolve antiphase brightness modulations (1 Hz) in abutting cortical representations. On the basis of these considerations, we suggest that our model can be an important tool in the design of appropriate stimuli for fMRI experiments on perceptual filling-in. For example, on the basis of simulations, we know that to empirically demonstrate filling-in of a gray figure by surrounding dynamic texture with fMRI, an active interpolation signal must exceed the passive neural spread of fMRI activity from the texture into the cortical representation of the figure. Our model suggests that this problem may be avoided only when using very large figure sizes. Overlay of simulated fMRI activity distributions in sheets of units representing boundary- and surface-processing systems in a M-area that respects retinotopy and cortical magnification can be helpful to precisely assess the possibility of discerning activity specifically related to perceptual filling-in. In our model, coming to this assessment is facilitated by NB-links that can map simulated activity distributions onto a flat map of a human subject's cortex in which empirical data can later be collected.

Insights from Modeling Recurrent Connections Between M-V2 and M-V4 The recurrent neural network model may offer a new account of modal and amodal completion/filling-in. First, the implementation of a recurrent loop between surface modules in M-V2 and M-V4 (see the second part of the chapter) is important for understanding what defines *modal* ("real") surface percepts. We suggest that modal completion or filling-in of a figure by its background requires spiking activity in figure representations in V2 and V4 comparable to the activity that would be observed in the absence of a figure on the background. Therefore only when active spreading occurs in V2 and V4 via spiking activity, the system has made a "decision" to consider the internally represented surface as "real." In this state, spiking neurons in V2 and spiking neurons in higher visual areas (e.g., V4) form functional circuits jointly representing the content of the filled-in percept. Although we only modeled spreading in the Troxler paradigm, we suggest that a "spreading" of the "real/visible" status of a surface also takes place during normal perception when there is a direct physical basis for the perceived surface, in which case recurrent processing might modulate bottom-up surface-related activity and the associated surface percept. Our discussion of modal completion/filling-in is in line with theories of conscious

visual experience that assign a crucial role to recurrent loops between lower- and higher-order visual areas (e.g., Lamme & Roelfsema, 2000; Tong, 2003). A successful recurrent loop might be signaled via synchronous activity between V2 and V4 neurons (Singer, 1999; Goebel, Muckli, & Kim, 2003). The strength of synchronous activity across the hierarchy would be determined by the amount of matching information that is exchanged between V2 and V4 neurons at corresponding topological regions.

Second, the implementation of a recurrent loop between surface modules in M-V2 and M-V4 (see the second part of the chapter) is also relevant for the interpretation of *amodal* completion. We suggest that the neural substrate of amodal completion is nothing else than the subthreshold activity related to parts of occluded surfaces that spreads within boundary and surface systems and thereby partially represents the invisible parts of the occluded surfaces. Because this "passive" spread of subthreshold activity is not driven by any physical input and because the spread within V2 can be expected to be more limited than that in V4, we suggest that there is insufficient coupling of topologically corresponding units in V2 and V4 to make the occluded surfaces "visible." The presence of a subthreshold "trace" that amodally completes occluded shapes may, however, interact with higher levels in the visual system when high-level shape or object representations are activated. This could explain why performance in some cognitive tasks (e.g., search) can be strongly modulated by manipulations that permit or prevent amodal completion (for a review, see Davis & Driver, 2003).

Limitations and Future Research Directions The model in its current state is sufficient to illustrate that computational modeling studies can provide important contributions to understanding the outcome of fMRI and neuronal recording experiments on the topic of perceptual filling-in and that they may help with the choice of appropriate stimuli and the invention of more sensitive experimental designs. However, an important limitation of the current model is that the connection weights in the model are fixed, and as a consequence, there is currently no modeling possible of attention or learning-induced plastic changes in connectivity. De Weerd and Pessoa (2003) have suggested that perceptual filling-in of a figure by a background may be comparable to the initial filling-in of a retinal scotoma and that it could form the beginning of a chain of plastic events leading to cortical remapping, as has been documented in visual and other sensory systems (e.g., Kaas et al., 2003). The implementation of learning rules might permit a simulation of these plastic effects.

acknowledgments We would like to thank David Janssen and Bert Jans for help in creating the network model as well as for many fruitful discussions.

REFERENCES

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso, & P. U. Tse (Eds.), *Visual perception, Part I, Fundamentals of vision: Low and mid-level processes in perception* (*Progress in brain research*, Vol. 154, pp. 93–120). Amsterdam: Elsevier.

Angelucci, A., Levitt, J. B., Walton, E. J., Hupe, J. M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *J. Neurosci.*, *22*(19), 8633–8646.

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nat. Neurosci.*, *10*, 1313–1321.

Arrington, K. F. (1994). The temporal dynamics of brightness filling-in. *Vis. Res.*, *34*, 3371–3387.

Bartfeld, E., & Grinvald A. (1992). Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. *Proc. Natl. Acad. Sci. USA*, *89*(24), 11905–11909.

Bender, M. B., & Teuber, L. H. (1946). Phenomena of fluctuation, extinction and completion in visual perception. *Arch. Neurol. Psychiatry (Chicago)*, *55*, 627–658.

Biederman, I. (2000). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vis.*, *13*, 241–253.

Blakeslee, B., & McCourt, M. E. (2008). Nearly instantaneous brightness indiction. *J. Vis.*, *8*(2)*:15*, 1–8.

Blakeslee, B., Pasieka, W., & McCourt, M. E. (2005). Oriented multiscale spatial filtering and contrast normalization: A parsimonious model for brightness induction in a continuum of stimuli including White, Howe and simultaneous brightness contrast. *Vis. Res.*, *45*, 607–615.

Blasdel, G. G. (1992). Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *J. Neurosci.*, *12*(8), 3115–3138.

Blasdel, G. G., & Lund, J. S. (1983). Termination of afferent axons in macaque striate cortex. *J. Neurosci.*, *3*(7), 1389–1413.

Blasdel, G. G., Lund, J. S., & Fitzpatrick, D. (1985). Intrinsic connections of macaque striate cortex: Axonal projections of cells outside lamina 4C. *J. Neurosci.*, *5*(12), 3350–3369.

Blasdel, G., Obermayer, K., & Kiorpes, L. (1995). Organization of ocular dominance and orientation columns in the striate cortex of neonatal macaque monkeys. *Vis. Neurosci.*, *12*(3), 589–603.

Bonhoeffer, T., Kim, D. S., Malonek, D., Shoham, D., & Grinvald, A. (1995). Optical imaging of the layout of functional domains in area 17 and across the area 17/18 border in cat visual cortex. *Eur. J. Neurosci.*, *7*(9), 1973–1988.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.*, *16*, 4207–4221.

Buckley, D., Frisby, J. P., & Mayhew, J. E. (1989). Integration of stereo and texture cues in the formation of discontinuities during three-dimensional surface interpolation. *Perception*, *18*(5), 563–588.

Clarke, F. J. J. (1957). Rapid light adaptation of localized areas of the extra-foveal retina. *Optica Acta*, *4*, 69–77.

Coppola, D., & Purves, D. (1996). The extraordinarily rapid disappearance of entopic images. *Proc. Natl. Acad. Sci. USA*, *3*, 8001–8004.

Cornelissen, F. W., Wade, A. R., Vladusich, T., Dougherty, R. F., & Wandell, B. A. (2006). No functional magnetic resonance imaging evidence for brightness and color filling-in in early human visual cortex. *J. Neurosci.*, *26*, 3634–3641.

Dakin, S. C., & Bex, B. J. (2003). Natural image statistics mediate brightness "filling in." *Proc. Biol. Sci.*, *270*(1531), 2341–2348.

Davey, M. P., Maddess, T., & Srinivasan, M. V. (1998). The spatiotemporal properties of the Craik-O'Brien-Cornsweet effect are consistent with "filling-in." *Vis. Res.*, *38*(13), 2037–2046.

Davis, G., & Driver, J. (2003). Effects of modal and amodal completion upon visual attention: A function for filling-in? In L. Pessoa & P. De Weerd (Eds.), *Filling-in: From perceptual completion to skill learning* (pp. 128–150). Oxford, UK: Oxford University Press.

De Weerd, P. (2006). Perceptual filling-in: More than the eye can see. In Martinez-Conde, S., Macknik, S. L., Martinez, L. M., Alonso, J. M., & Tse, P. U. (Eds.), *Visual perception, Part I, Fundamentals of vision: Low and mid-level processes in perception* (*Progress in brain research*, Vol. 154, pp. 227–245). Amsterdam: Elsevier.

De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Perceptual filling-in: A parametric study. *Vis. Res.*, *38*, 2721–2734.

De Weerd, P., Gattass, R., Desimone, R., & Ungerleider, L. G. (1995). Responses of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma. *Nature*, *377*, 731–734.

De Weerd, P., Karni, A., Kastner, Ungerleider, L. G., & Jezzard, P. (1997). *An investigation of fMRI resolution in the visual cortex*. Paper presented at the Third International Conference on Functional Mapping of the Human Brain.

De Weerd, P., & Pessoa, L. (2003). Filling-in: More than meets the eye. In L. Pessoa & P. De Weerd (Eds.), *Filling-in: From perceptual completion to skill learning* (pp. 295–322). Oxford, UK: Oxford University Press.

Denett, D. (1991). *Consciousness explained*. Boston: Little Brown.

Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *J. Neurophysiol.*, *57*(3), 835–868.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 2, pp. 267–299). Amsterdam: Elsevier.

DeYoe, E. A., & van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends Neurosci.*, *11*, 219–226.

Ditchburn, R. W., & Ginsborg, B. L. (1953). Involuntary eye movements during fixation. *J. Physiol.*, *118*, 1–17.

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex*, *7*(2), 181–192.

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., & Chichilnisky, E. J., et al. (1994). fMRI of human visual cortex. *Nature*, *369*(6481), 525.

Farias, M. F., Gattass, R., Piñón, M. C., & Ungerleider, L. G. (1997). Tangential distribution of cytochrome oxidase-rich blobs in the primary visual cortex of macaque monkeys. *J. Comp. Neurol.*, *386*(2), 217–228.

Fiorani, M., Rosa, M. G. P., Gattass, R., & Rocha-Miranda, C. E. (1992). Dynamic surrounds of receptive fields in primate striate cortex: A physiological basis for perceptual completion. *Proc. Natl. Acad. Sci. USA*, *89*, 8547–8551.

Fitzpatrick, D., Lund, J. S., & Blasdel, G. G. (1985). Intrinsic connections of macaque striate cortex: Afferent and efferent connections of lamina 4C. *J. Neurosci.*, *5*(12), 3329–3349.

Friedman, H. S., Zhu, H., & von der Heydt, R. (1999). Color filling-in under steady fixation conditions: Behavioral demonstration in monkeys and humans. *Perception*, *28*, 1383–1395.

Friedman, H. S., Zhou, H., & von der Heydt, R. (2002). The coding of uniform color figures in monkey visual cortex. *J. Physiol.*, *548*(2), 593–613.

Frisby, J. P., Buckley, D., Wishart, K. A., Porrill, J., Gårding, J., & Mayhew J. E. (1995). Interaction of stereo and texture cues in the perception of three-dimensional steps. *Vis. Res.*, *35*(10), 1463–1472.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*, 30–40.

Fujita, M. (1993). Color filling-in in foveal vision. *Soc. Neurosci. [Abstracts]*, *19*(3), 1802.

Gattass, R., Gross, C. G., & Sandell, J. H. (1981). Visual topography of V2 in the macaque. *J. Comp. Neurol.*, *21*, 519–539.

Gattass, R., Sousa, A. P. B., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *J. Neurosci.*, *8*(8), 1831–1845.

Gerrits, H. J. M., de Haan, B., & Vendrik, A. J. H. (1966). Experiments with retinal stabilized images: Relations between the observations and neural data. *Vis. Res.*, *6*, 427–440.

Gerrits, H. J. M., & Vendrik, A. J. H. (1970). Simultaneous contrast, filling-in process and information processing in man's visual system. *Exp. Brain Res.*, *11*, 411–440.

Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic cortico-cortical connections in cat visual cortex. *J. Neurosci.*, *9*, 2432–2442.

Goebel, R. (1993). Perceiving complex visual scenes: An oscillator neural network model that integrates selective attention, perceptual organisation, and invariant recognition. In C. L. Giles, S. J. Hanson, & J. D. Cowan (Eds.), *Advances in neural information processing systems 5* (pp. 903–910). San Mateo, CA: Morgan Kaufmann.

Goebel, R., & Horwitz, B. (2008). Brain imaging data and large-scale neural network models: Tightening the link. In preparation.

Goebel, R., Muckli, L., & Kim, D.-S. (2003). The visual system. In G. Paxinos & J. K. Mai (Eds.), *The human nervous system* (2nd ed., pp. 1280–1305). New York: Academic Press.

Gove, A., Grossberg, S., & Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Vis. Neurosci.*, *12*, 1027–1052.

Grimson, W. E. (1982). A computational theory of visual surface interpolation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *298*(1092), 395–427.

Grossberg, S. (1987a). Cortical dynamics of three-dimensional form, color, and brightness perception: I. Monocular theory. *Percept. Psychophys.*, *41*, 87–116.

Grossberg, S. (1987b). Cortical dynamics of three-dimensional form, color, and brightness perception: II. Binocular theory. *Percept. Psychophys.*, *41*, 117–158.

Grossberg, S. (1997). Cortical dynamics of 3D figure ground perception of 2D pictures. *Psychol. Rev.*, *104*, 618–658.

Grossberg, S. (2003a). Filling-in the forms: Surface and boundary interaction in visual cortex. In L. Pessoa & P. De Weerd (Eds.), *Filling-in: From perceptual completion to skill learning* (pp. 13–37). Oxford, UK: Oxford University Press.

Grossberg, S. (2003b). How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. *Behav. Cogn. Neurosci. Rev.*, *2*, 47–76.

Grossberg, S., & Raizada, R. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vis. Res.*, *40*, 1413–1432.

Grossberg, S., & Todorovic, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Percept. Psychophys.*, *43*, 241–277.

Hamburger, K., Prior, H., Sarris, V., & Spillmann, L. (2005). Filling-in with colour: Different modes of surface completion. *Vis. Res.*, *46*(6–7), 1129–1138.

Hanazawa, A., & Komatsu, H. (2001). Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J. Neurosci.*, *21*(12), 4490–4497.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

Heider, B., Meskenaite, V., & Peterhans, E. (2000). Anatomy and physiology of a neural mechanism defining depth order and contrast polarity at illusory contours. *Eur. J. Neurosci.*, *12*, 4117–4130.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *J. Vis.*, *4*(12), 967–992.

Hindi Attar, C., Hamburger, K., Rosenholtz, R., Götzl, H., & Spillmann, L. (2007). Uniform versus random orientation in fading and filling-in. *Vis. Res.*, *47*(24), 3041–3051.

Hsieh, P. J., & Tse, P. U. (2006). Illusory color mixing upon perceptual fading and filling-in does not result in "forbidden colors." *Vis. Res.*, *46*(14), 2251–2258.

Huang, X., & Paradiso, M. A. (2008). V1 response timing and surface filling-in. *J. Neurophysiol.*, *100*(1), 539–547.

Hubel, D. H., & Livingstone, M. S. (1987). Segregation of form, color, and stereopsis in primate area 18. *J. Neurosci.*, *7*(11), 3378–3415.

Hubel, D. H., & Livingstone, M. S. (1990). Color and contrast sensitivity in the lateral geniculate body and primary visual cortex of the macaque monkey. *J. Neurosci.*, *10*(7), 2223–2237.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *J. Physiol. Lond.*, *148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. Lond.*, *160*, 106–154.

Hughes, H. C., Nozawa, G., & Kitterle. F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *J. Cogn. Neurosci.*, *8*(3), 197–230.

Hung, C. P., Ramsden, B. M., Chen, L. M., & Roe, A. W. (2001). Building surfaces from borders in areas 17 and 18 of the cat. *Vis. Res.*, *41*(10–11), 1389–1407.

Hung, C. P., Ramsden, B. M., & Roe, A. W. (2007). A functional circuitry for edge-induced brightness perception. *Nat. Neurosci.*, *10*(9), 1185–1190.

Hupé, J. M., James, A. C., Girard, P., & Bullier, J. (2001). Response modulations by static texture surround in area V1 of the macaque monkey do not depend on feedback connections from V2. *J. Neurophysiol.*, *85*(1), 146–163.

Hupé, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *J. Neurophysiol.*, *85*(1), 134–145.

Kaas, J. H., Collins, C. E. & Chino, Y. M. (2003). The reactivation and reorganization of retinotopic maps in visual cortex after retinal and cortical lesions. In L. Pessoa & P. De Weerd (Eds.), *Filling-in: From perceptual completion to skill learning* (pp. 295–322). Oxford, UK: Oxford University Press.

Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *J. Neurosci.*, *23*, 3016–3027.

Kayama, Y., Riso, R. R., Bartlett, J. R., & Doty, R. W. (1979). Luxotonic responses of units in macaque striate cortex. *J. Neurophysiol.*, *42*(6), 1495–1517.

Kinoshita, M., & Komatsu, H. (2001). Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *J. Neurophysiol.*, *86*(5), 2559–2270.

Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nat. Rev. Neurosci.*, *7*, 220–231.

Komatsu, H., Kinoshita, M., & Murakami, I. (2000). Neural responses in the retinotopic representation of the blind spot in the macaque V1 to stimuli for perceptual filling-in. *J. Neurosci.*, *20*, 9310–9319.

Komatsu, H., & Murakami, I. (1994). Behavioral evidence of filling-in at the blind spot of the monkey. *Vis. Neurosci.*, *11*, 1103–1113.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.*, *23*(11), 571–579.

Landisman, C. E., & Ts'o, D. Y. (2002). Color processing in macaque striate cortex: Relationships to ocular dominance, cytochrome oxidase, and orientation. *J. Neurophysiol.*, *87*(6), 3126–3137.

Livingstone, M. S., & Hubel, D. H (1983). Specificity of cortico-cortical connections in monkey visual system. *Nature*, *304*(5926), 531–534.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annu. Rev. Physiol.*, *66*, 735–769.

Lu, H. D., & Roe, A. W. (2008). Functional organization of color domains in V1 and V2 of macaque monkey revealed by optical imaging. *Cereb. Cortex*, *18*(3), 516–533.

Lund, J. S., Yoshioka, T., & Levitt, J. B. (1993). Comparison of intrinsic connectivity in different areas of macaque monkey cerebral cortex. *Cereb. Cortex*, *3*, 148–162.

Maguire, W. M., & Baizer, J. S. (1982). Luminance coding of briefly presented stimuli in area 17 of the rhesus monkey. *J. Neurophysiol.*, *47*(1), 128–137.

Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F. Q. & Leopold, D. A. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat. Neurosci.*, *11*(10), 1193–1200.

Malach, R., Tootell, R. B., & Malonek, D. (1994). Relationship between orientation domains, cytochrome oxidase stripes, and intrinsic horizontal connections in squirrel monkey area V2. *Cereb. Cortex*, *4*(2), 151–165.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.

Martinez-Conde, S. (2006). Fixational eye movements in normal and pathological vision. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso, & P. U. Tse (Eds.), *Visual perception, Part I, Fundamentals of vision: Low and mid-level processes in perception* (*Progress in brain research*, Vol. 154, pp. 151–176). Amsterdam: Elsevier.

Matsumoto, M., & Komatsu, H. (2005). Neural responses in the macaque V1 to bar stimuli with various lengths presented on the blind spot. *J. Neurophysiol.*, *93*(5), 2374–2387.

McCourt, M. E. (1982). A spatial frequency dependent grating-induction effect. *Vis. Res.*, *22*, 119–134.

McGuire, B. A., Gilbert, C. D., Rivlin, P. K., & Wiesel, T. N. (1991). Targets of horizontal connections in macaque primary visual cortex. *J. Comp. Neurol.*, *305*, 370–392.

Mendola, J. D., Conner, I. P., Sharma, S., Bahekar, A., & Lemieux, S. (2006). fMRI measures of perceptual filling-in in the human visual cortex. *J. Cogn. Neurosci.*, *18*, 363–375.

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proc. R. Soc. Lond. B Biol. Sci.*, *235*(1280), 221–245.

Motoyoshi, I. (1999). Texture filling-in and texture segregation revealed by transient masking. *Vis. Res.*, *39*(7), 1285–1291.

Murakami, I., Komatsu, H., & Kinoshita, M. (1997). Perceptual filling-in at the scotoma following a monocular retinal lesion in the monkey. *Vis. Neurosci.*, *14*, 89–101.

Nakamura, H., Gattass, R., Desimone, R., & Ungerleider, L. G. (1993). The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. *J. Neurosci.*, *13*(9), 3681–3691.

Nealy, T. A., & Maunsell, J. H. (1994). Magnocellular and parvocellular contributions to the responses of neurons in macaque striate cortex. *J. Neurosci.*, *14*(4), 2069–2079.

Neumann, H., Pessoa, L., & Hansen, T. (2001). Visual filling-in for computing perceptual surface properties. *Biol. Cybern.*, *85*, 355–369.

Oeltermann, A., Augath, M. A., & Logothetis, N. K. (2007). Simultaneous recording of neuronal signals and functional NMR imaging. *Magn. Reson. Imaging*, *25*(6), 760–774.

Paradiso, M. A., & Hahn, S. (1996). Filling-in percepts produced by luminance modulation. *Vis. Res.*, *36*, 2657–2663.

Paradiso, M. A., & Nakayama, K. (1991). Brightness perception and filling-in. *Vis. Res.*, *31*, 1221–1236.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nat. Neurosci.*, *5*(12), 1332–1338.

Perna, A., Tosetti, M., Montanaro, D., & Morrone, M. C. (2005). Neuronal mechanisms for illusory brightness perception in humans. *Neuron*, *47*, 645–651.

Pessoa, L., & Neumann, H. (1998). Why does the brain fill-in? *Trends Cogn. Sci.*, *2*, 422–424.

Pessoa, L., Thompson, E., & Noë, A. (1998). Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception. *Behav. Brain Sci.*, *21*, 723–748.

Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex: II. Contours bridging gaps. *J. Neurosci.*, *9*, 1749–1763.

Purves, D., Shimpi, A., & Lotto, R. B. (1999). An empirical explanation of the cornsweet effect. *J. Neurosci.*, *19*(19), 8542–8551.

Qiu, F. T., & von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. *Neuron*, *47*(1), 155–166.

Raizada, R., & Grossberg, S. (2001). Context-sensitive bindings by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cogn.*, *8*, 431–466.

Ramachandran, V. S., & Gregory, R. L. (1991). Perceptual filling-in of artificially induced scotomas in human vision. *Nature*, *350*, 699–702.

Ramachandran, V. S., Gregory, R. L., & Aiken, W. (1993). Perceptual fading of visual texture borders. *Vis. Res.*, *33*, 717–721.

Riggs, L. A., & Ratliff, F. (1952). The effects of counteracting the normal movements of the eye. *J. Opt. Soc. Am. [A]*, *42*, 872–873.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *J. Opt. Soc. Am. [A]*, *43*, 495–501.

Roe, A. W., Lu, H. D., & Hung, C. P. (2005). Cortical processing of a brightness illusion. *Proc. Natl. Acad. Sci. USA*, *102*, 3869–3874.

Roe, A. W., & Ts'o, D. Y. (1995). Visual topography in primate V2: Multiple representations across functional stripes. *J. Neurosci.*, *15*, 3689–3715.

Roe, A. W., & Ts'o, D. Y. (1999). Specificity of color connectivity between primate V1 and V2. *J. Neurophysiol.*, *82*(5), 2719–2730.

Rossi, A. F., & Paradiso, M. A. (1996). Temporal limits of brightness induction and mechanisms of brightness perception. *Vis. Res.*, *36*, 1391–1398.

Rossi, A. F., & Paradiso, M. A. (1999). Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex. *J. Neurosci.*, *19*, 6145–6156.

Sakaguchi, Y. (2001). Target/surround asymmetry in perceptual filling-in. *Vis. Res.*, *41*(16), 2065–2077.

Sakaguchi, Y. (2006). Contrast dependency in perceptual filling-in. *Vis. Res.*, *46*(20), 3304–3312.

Salmela, V. R., & Laurinen, P. I. (2007). Brightness processing in the visual cortex. *Neurosci. Lett.*, *420*(2), 160–162.

Sasaki, Y., & Watanabe, T. (2004). The primary visual cortex fills in color. *Proc. Natl. Acad. Sci. USA*, *101*, 18251–18256.

Schein, S. J., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *J. Neurosci.*, *10*(10), 3369–3389.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893.

Sergent, J. (1988). An investigation into perceptual completion in blind areas of the visual field. *Brain*, *111*, 347–373.

Shipp, S., & Zeki, S. (2002). The functional organization of area V2: I. specialization across stripes and layers. *Vis. Neurosci.*, *19*(2), 187–210.

Shoham, D., Hübener, M., Schulze, S., Grinvald, A., & Bonhoeffer, T. (1997). Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex. *Nature*, *385*(6616), 529–533.

Sincich, L. C., & Horton, J. C. (2005). The circuitry of V1 and V2: Integration of color, form, and motion. *Annu. Rev. Neurosci.*, *28*, 303–326.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*, 49–65.

Smirnakis, S. M., Brewer, A. A., Schmid, M. C., Tolias, A. S., Schüz, A., Augath, M., et al. (2005). Lack of long-term cortical reorganization after macaque retinal lesions. *Nature*, *435*(7040), 300–307.

Spillmann, L., & De Weerd, P. (2003). Mechanisms of surface completion: Perceptual filling-in of texture. In L. Pessoa & P. De Weerd (Eds.), *Filling-in: From perceptual completion to skill learning* (pp. 295–322). Oxford, UK: Oxford University Press.

Spillmann, L., & Kurtenbach, A. (1992). Dynamic noise backgrounds facilitate target fading. *Vis. Res.*, *32*, 1941–1946.

Stromeyer, C. F., III., Kronauer, R. E., Madsen, J. C., & Klein, S. A. (1984). Opponent-movement mechanisms in human vision. *J. Opt. Soc. Am. [A]*, *1*, 876–884.

Stürzel, F., & Spillmann, L. (2001). Texture fading correlates with stimulus salience. *Vis. Res.*, *41*(23), 2969–2977.

Tanabe, S., Doi, T., Umeda, K., & Fujita, I. (2005). Disparity-tuning characteristics of neuronal responses to dynamic random-dot stereograms in macaque visual area V4. *J. Neurophysiol.*, *94*(4), 2683–2699.

Todorovic, D. (1987). The Craik-O'Brien-Cornsweet effect: New varieties and their theoretical implications. *Percept. Psychophys.*, *42*, 545–560.

Tolias, A. S., Smirnakis, S. M., Augath, M. A., Trinath, T., & Logothetis, N. K. (2001). Motion processing in the macaque: Revisited with functional magnetic resonance imaging. *J. Neurosci.*, *21*(21), 8594–8601.

Tong, F. (2003). Primary visual cortex and visual awareness. *Nat. Rev. Neurosci.*, *4*, 219–229.

Tremere, L., Pinaud, R., & De Weerd, P. (2003). Contributions of inhibitory mechanisms to perceptual completion and cortical reorganization. In L. Pessoa & P. De Weerd (Eds.), *Filling-in: From perceptual completion to skill learning* (pp. 295–322). Oxford, UK: Oxford University Press.

Treue, S., Andersen, R. A., Ando, H., & Hildreth, E. C. (1995). Structure-from-motion: Perceptual evidence for surface interpolation. *Vis. Res.*, *35*(1), 139–148.

Troxler, D. (1804). Über das Verschwinden gegebener Gegenstände innerhalb unseres Gesichtskreises. In K. Himly & J. A. Schmidt (Eds.), *Ophthalmolgische bibliothek* (Vol. 2, pp. 1–119). Jena: Fromann.

Ts'o, D. Y., & Gilbert, C. D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *J. Neurosci.*, *8*, 1712–1727.

Ts'o, D. Y., Roe, A. W., & Gilbert, C. D. (2001). A hierarchy of the functional organization for color, form and disparity in primate visual area V2. *Vis. Res.*, *41*(10–11), 1333–1349.

von der Heydt, R., Friedman, F. S., & Zhou, H. (2003). Seaching for the neural mechanism of color filling-in. In L. Pessoa & P. De Weerd (Eds.), *Filling-in: From perceptual completion to skill learning* (pp. 295–322). Oxford, UK: Oxford University Press.

von der Heydt, R., & Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex: I. Lines of pattern discontinuity. *J. Neurosci.*, *9*, 1731–1748.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, *224*, 1260–1262.

Walls, G. K. (1954). The filling-in process. *Am. J. Optom. Arch. Am. Acad. Optom.*, *31*, 239–241.

Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*, *272*, 1665–1668.

Wang, Y., Xiao, Y., & Felleman, D. J. (2007). V2 thin stripes contain spatially organized representations of achromatic luminance change. *Cereb. Cortex*, *17*(1), 116–129.

Weil, R. S., Watkins, S., & Rees, G. (2008). Neural correlates of perceptual completion of an artificial scotoma in human visual cortex measured using functional MRI. *NeuroImage*, *42*(4), 1519–1528.

Welchman, A. E., & Harris, J. M. (2001). Filling-in the details on perceptual fading. *Vis. Res.*, *41*(16), 2107–2117.

Welchman, A. E., & Harris, J. M. (2003). Is neural filling-in necessary to explain the perceptual completion of motion and depth information? *Proc. R. Soc. Lond. B Biol. Sci.*, *270*(1510), 83–90.

Xiao, Y., & Felleman, D. J. (2004). Projections from primary visual cortex to cytochrome oxidase thin stripes and interstripes of macaque visual area 2. *Proc. Natl. Acad. Sci. USA*, *101*(18), 7147–7151.

Xiao, Y., Zych, A., & Felleman, D. J. (1999). Segregation and convergence of functionally defined V2 thin stripe and interstripe compartment projections to area V4 of macaques. *Cereb. Cortex*, *9*(8), 792–804.

Yabuta, N. H., & Callaway, E. M. (1998a). Functional streams and local connections of layer 4C neurons in primary visual cortex of the macaque monkey. *J. Neurosci.*, *18*(22), 9489–9499.

Yabuta, N. H., & Callaway, E. M. (1998b). Cytochrome-oxidase blobs and intrinsic horizontal connections of layer 2/3 pyramidal neurons in primate V1. *Vis. Neurosci.*, *15*(6), 1007–1027.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press.

Yokota, M., & Yokota, Y. (2004). Spatio-temporal frequency characteristics of perceptual filling-in. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, *1*, 639–642.

Yokota, M., & Yokota, Y. (2005). Facilitation of perceptual filling-in for spatio-temporal frequency of dynamic textures. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, *3*, 2926–2931.

Yoshioka, T., Blasdel, G. G., Levitt, J. B., & Lund, J. S. (1996). Relation between patterns of intrinsic lateral connectivity, ocular dominance, and cytochrome oxidase-reactive regions in macaque monkey striate cortex. *Cereb. Cortex*, *6*(2), 297–310.

Yoshioka, T., & Dow, B. M. (1996). Color, orientation and cytochrome oxidase reactivity in areas V1, V2 and V4 of macaque monkey visual cortex. *Behav. Brain Res.*, *76*(1–2), 71–88.

Zhang, X., & von der Heydt, R. (1995). Determinants of filling-in. *Invest. Ophthalmol. Visual Sci.*, *36*, S469.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.*, *20*, 6594–6611.

31 Neural Transformation of Object Information by Ventral Pathway Visual Cortex

CHARLES E. CONNOR, ANITHA PASUPATHY, SCOTT BRINCAT, AND YUKAKO YAMANE

ABSTRACT Object perception is a critical aspect of cognition, and is essential to understanding the world we inhabit. It is also one of the brain's most remarkable computational abilities, considering the enormously complex, variable mapping between retinal images and physical objects. Retinal images are transformed into mental representations of objects by the ventral pathway of visual cortex. The vast dimensionality of the retinal image is compressed into a compact representation of object part configurations. This explicit representation of configural structure may serve as the basis for recognition, evaluation, and physical interaction with objects.

We live in a world of objects, and that world is familiar and comprehensible only because we are so good at recognizing and understanding those objects. Object perception is computationally difficult because of the high dimensionality of the retinal input (on the order of 10^6 channels) and the extreme variability in input patterns produced by any given object (depending on position, distance, orientation, lighting, etc.). The brain must transform this complex, variable retinal input into compact, stable representations of useful object information. This transformation is carried out by the ventral pathway of visual cortex (Ungerleider & Mishkin, 1982; Felleman & Van Essen, 1991), which splits off from the rest of the visual hierarchy at the connection between areas V2 and V4. Beyond V4, object information is processed through a posterior-to-anterior series of stages in inferior occipital and temporal cortex.

The first-order question about the ventral pathway transformation is how object information is encoded at each stage. This chapter describes the current understanding of object shape coding in area V4 and inferotemporal cortex (IT) based on neurophysiological studies in macaque monkeys. The second-order and more difficult question is what computational mechanisms underlie the transformations between coding stages. Preliminary analyses of recurrent network mechanisms supporting the V4-to-IT transformation are described here.

CHARLES E. CONNOR Johns Hopkins University, Baltimore, Maryland
ANITHA PASUPATHY University of Washington, Seattle, Washington
SCOTT BRINCAT Massachusetts Institute of Technology, Boston, Massachusetts
YUKAKO YAMANE Riken Brain Science Institute, Saitama, Japan

Retinal signals must be transformed to support object vision

Two factors make the original retinal representation of objects unsuitable to support object perception. One factor is high dimensionality. The retinal representation is essentially a megapixel spatial map of local contrast that replicates the form of the optical image. A million-dimensional signal cannot be directly accessed by other brain regions to guide behavior and cannot be stored in memory. This high-dimensional pixel map must be transformed into a tractable, explicit code for useful object information. As will be described below, the ventral pathway achieves this by recoding large regions of the pixel map as object boundary fragments characterized by geometric derivatives. Entire objects are represented as spatial configurations of boundary fragments.

The other factor is variability. Any given object produces a potentially infinite range of retinal input patterns. This is due to the continually shifting relationship between the input spatial reference frame of the eye and the signal source reference frame of the object. The variable mapping between eye images and objects makes the retinal representation far too unstable for cognitive access and memory storage. As will be described below, the ventral pathway derives a more stable representation by transforming spatial information from eye coordinates into a reference frame that is at least partially defined by the object itself.

Object boundary fragments are summarized by geometric derivatives

The retinal response patterns produced by natural images are far from random. (Most random pixel patterns look like

television snow.) They are dominated by local correlations that are determined by the structure of objects in our world, which in turn reflect the constraints of physics, material properties, biological growth processes, and artifactual construction. Because of these constraints, object boundaries are relatively smooth and continuous on a local level, producing smooth, continuous contrast boundaries in the retinal image. This creates an opportunity for massive compression of the pixel map representation. The highly correlated pixel values along regions of smooth, continuous contrast can be redescribed in terms of contrast boundary derivatives. For example, an image region that contains a long, vertical contrast edge comprises many pixel values but can be redescribed with a single slope or orientation value. The visual system exploits this opportunity in primary visual cortex (V1) with neurons that are tuned for orientation (and spatial frequency) of local contrast regions (Hubel & Wiesel, 1959, 1965, 1968). These tuning functions provide a basis set for representing local orientation. Every point in visual space is represented by V1 neurons with a range of orientation-tuning peaks, and the local hill of activity among these neurons encodes local contrast orientation. Computational studies have demonstrated that this is an optimal scheme for compressing fragments of natural images (Olshausen & Field, 1996; Vinje & Gallant, 2000).

At successive processing stages in visual cortex, further compression is achieved by neurons with progressively larger receptive fields (RFs) that summarize larger image regions. In parafoveal V4, RFs encompass several degrees of visual angle (Gattas, Sousa, & Gross, 1988). On this larger scale, contrast boundaries more frequently undergo orientation changes within the RF and therefore can no longer be effectively summarized with a single first-order derivative. However, owing to the continuity and relative smoothness of natural objects, these larger boundary regions can be summarized with a combination of first- and second-order derivatives. Orientation often changes at a relatively constant rate that can be represented with a single second-order derivative: curvature. Convex or positive curvature (protrusion of the boundary away from the object interior) ranges from near zero (flat, infinite radius) to shallow to sharp. At the limit of sharpness, curvature becomes infinite (zero radius) and is perceived as a discontinuity in orientation (a point or angle). Concave curvature (indentation into the object interior) likewise ranges from shallow to infinite.

Visual cortex exploits this larger-scale structural regularity by explicitly representing curvature and orientation of contrast boundary fragments in area V4. Just as V1 encodes tiny boundary fragments with a basis set of orientation tuning functions, V4 encodes larger boundary fragments with a basis set of tuning functions in the curvature/orientation domain. The tuning domain has been sampled in V4 neural recording experiments with the kind of stimuli plotted in figure 31.1*A* (Pasupathy & Connor, 1999). Average neural responses of three V4 neurons are plotted in figures 31.1*B*–31.1*D*. Darker backgrounds correspond to higher response rates (see gray-scale). The neurons in figures 31.1*B* and 31.1*C* exemplify tuning in the orientation/curvature domain. The figure 31.1*B* neuron responds to sharp convex curvature oriented (pointing) toward the left and upper left. The figure 31.1*C* neuron responds to shallow convex curvature oriented toward the right and lower right. Figure 31.1*D* exemplifies V4 neurons tuned for orientation at zero curvature, responding to all stimuli with a nearly flat component at the preferred orientation. Curvature/orientation tuning reflects a further compression of object boundary information. Larger boundary fragments that were originally represented by many component orientation signals in V1 can be summarized with a single hill of activity in the V4 curvature/orientation domain.

Objects are represented as spatial configurations of boundary fragments

The boundary fragments encoded by V4 neurons are large enough to constitute a basis set for structural or parts-based shape representation. According to structural shape-coding theories (Biederman, 1987; Marr & Nishihara, 1978; Milner, 1974; Selfridge, 1959; Sutherland, 1968), objects are represented as spatial configurations of common parts. Structural coding schemes have several advantages. First, they have a comparatively low dimensionality; an object that was originally represented by 10^4–10^6 pixels could be redescribed as a configuration of parts numbering in the 10^1–10^2 range. Second, structural coding is highly generative, owing to the combinatorial explosion of part configurations, and therefore has the capacity for representing a virtual infinity of objects with a finite set of neural signals. Third, to the extent to which the spatial reference frame is defined by the object itself, structural representations can be stable across views. Fourth, explicit structural information could be used not only for recognition, but also for physical evaluation and guidance of physical interactions with objects. Structural coding also seems consistent with our linguistic tendency to describe objects as configurations of parts.

These theoretical considerations and the figure 31.1 neurophysiological result suggest that V4 instantiates a structural object representation based on boundary fragments. A critical prediction of this hypothesis is that neurons maintain parts-level selectivity across different global shape contexts. This prediction conflicts with the standard, intuitive notion that ventral pathway neurons are selective for a single range of stimulus shapes, centered on a "best stimulus" that evokes the strongest response and defines the information conveyed by that neuron. According to structural theories, the same neuron should respond strongly to an infinite variety of stimulus shapes as long as they contain the parts-level

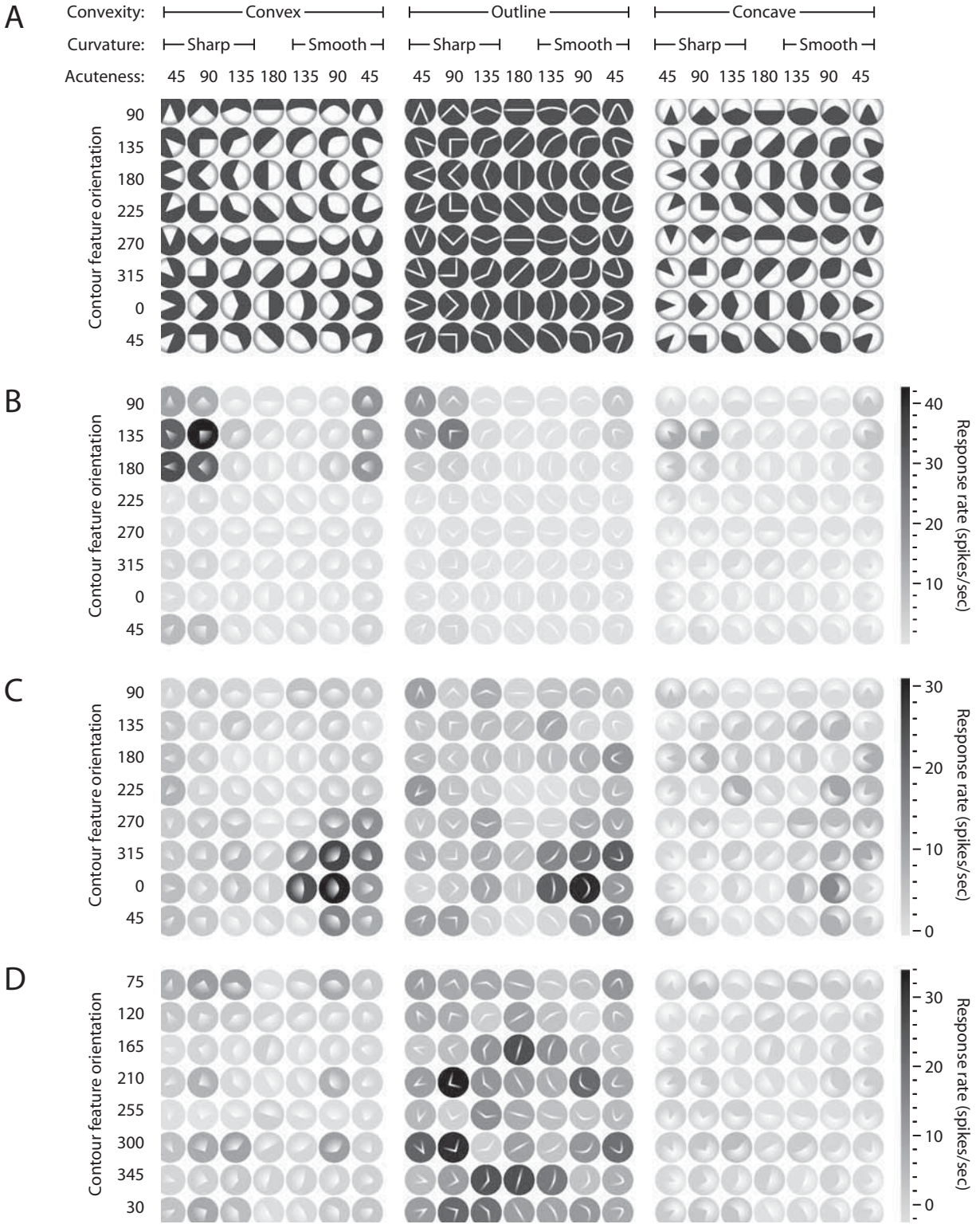


FIGURE 31.1 V4 neural tuning in the orientation/curvature domain. (*A*) Stimulus set comprising multiple levels of convex, outline, and concave boundary curvature at eight orientations. Spike activity of well-isolated individual neurons was recorded from lower visual field representations in V4 of rhesus macaque monkeys performing a fixation task to stabilize eye position. Stimuli were flashed in the neuron's RF for 500 ms each in random order. The stimulus was at full illumination within the RF perimeter and gradually faded to the background color outside the RF (only part of the fading is shown in these stimulus icons; the circular boundaries were not part of the stimuli). (*B*) Average response (across five repetitions) of a V4 neuron to each stimulus is indicated by background color behind each stimulus (black corresponds to 40 spikes per second; see the scale bar at right). This neuron is tuned along both the curvature dimension (for sharp convexity) and the orientation dimension (around 135 degrees, which here means sharp convexities pointing toward the upper left). (*C*) Average responses of another V4 neuron tuned for shallow curvature oriented toward the right (0 degrees). (*D*) Average responses of another V4 neuron tuned for orientation and zero curvature, like most lower-level neurons. (From Pasupathy & Connor (1999). Used with permission from *Journal of Neurophysiology*.)

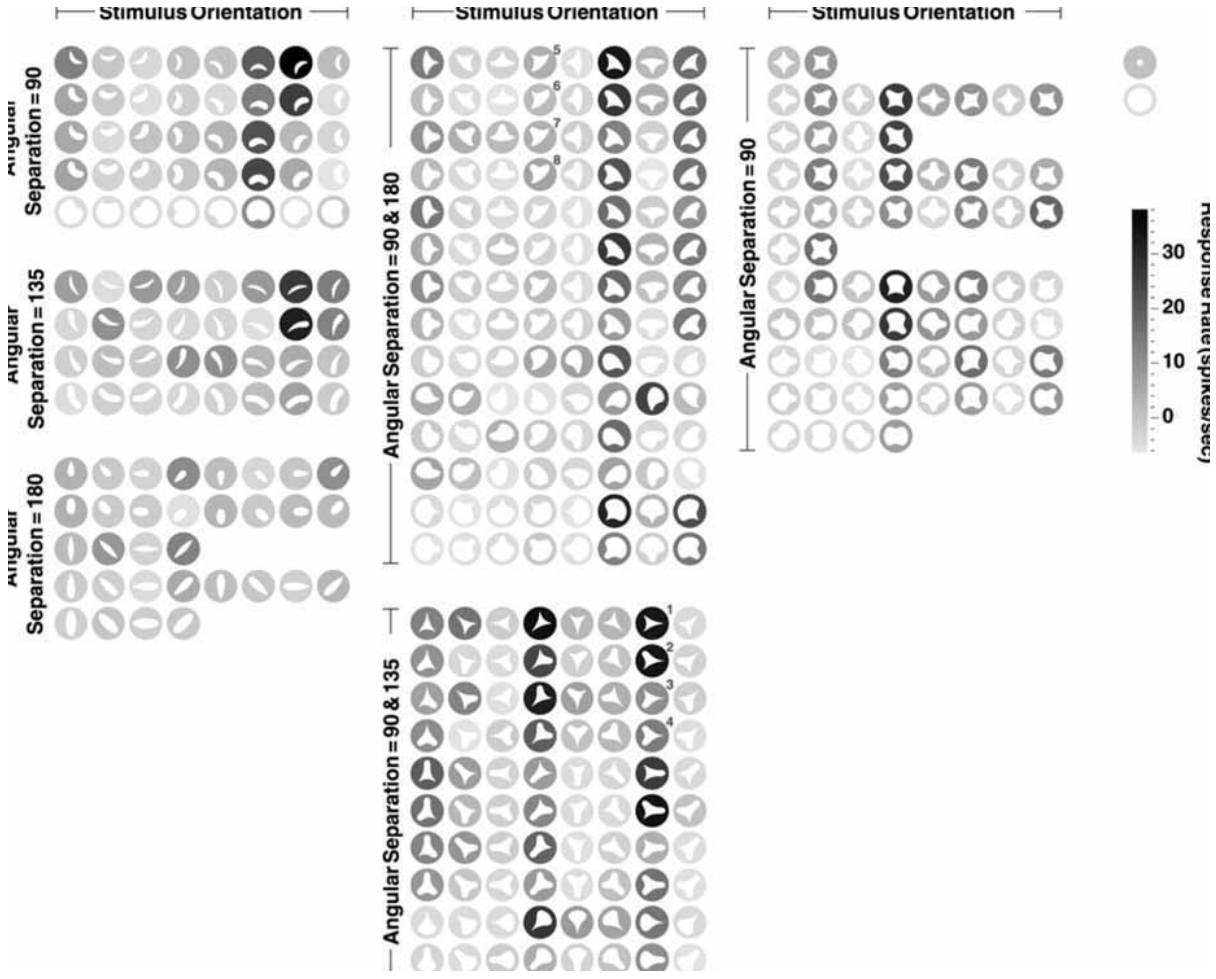


FIGURE 31.2 Responses of a V4 neuron to stimuli constructed by factorial combination of boundary fragments. Average responses across five repetitions are indicated by the background color for each stimulus. (The circular background was not part of the display.) This neuron was tuned for boundary fragments consisting of shallow concave curvature facing downward (270 degrees) adjacent to sharp convex curvature facing to the lower left (225 degrees). This tuning for local boundary structure remained consistent across wide variations in global stimulus shape. (From Pasupathy & Connor (2001). Used with permission from *Journal of Neurophysiology*.)

structure encoded by that neuron. This counterintuitive prediction has been confirmed in experiments exemplified by figure 31.2 (Pasupathy & Connor, 2001). For these experiments, a large set of stimuli was constructed by factorial combination of boundary fragments, so any given fragment appeared in a diverse set of global shapes. Each stimulus was presented entirely within the V4 neuron's RF, providing a strong test of consistency across global shape, since other parts of the shape could directly influence responses. This neuron responded to shapes that contained broad concave curvature facing downward adjacent to sharp convexities pointing to the lower left (see the stimuli labeled 1 and 2; either feature alone evoked weaker responses, as in stimuli 3–8). As can be seen in figure 31.2, this local structure evoked strong responses across wide variations in global stimulus shape. Thus V4 neurons encode parts-level, not global, shape. Analysis at the neural population level has demonstrated that these boundary fragment signals could support complete structural shape representations (Pasupathy & Connor, 2002).

Configural coding of object structure becomes even more explicit at the next processing stage in posterior IT (PIT). Neurons in PIT are tuned for spatial configurations of multiple V4-like boundary fragments (Brincat & Connor, 2004). The stimulus responses of the figure 31.3 example neuron (figures 31.3*A* and 31.3*B*) reflect combined tuning for the boundary fragments diagrammed in the figure 31.3*C* tuning model. The best-fit model for this cell comprised tuning for two sharp concavities (labeled A and B) oriented toward the lower left and lower right, respectively, and positioned to the left of object center, combined with flat or shallow curvature facing to the right and positioned to the right of object center (labeled C). The model includes an inhibitory term corresponding to the concavity labeled D. Figure 31.3*D* illustrates how these boundary fragment sensitivities interact to determine neural responses.

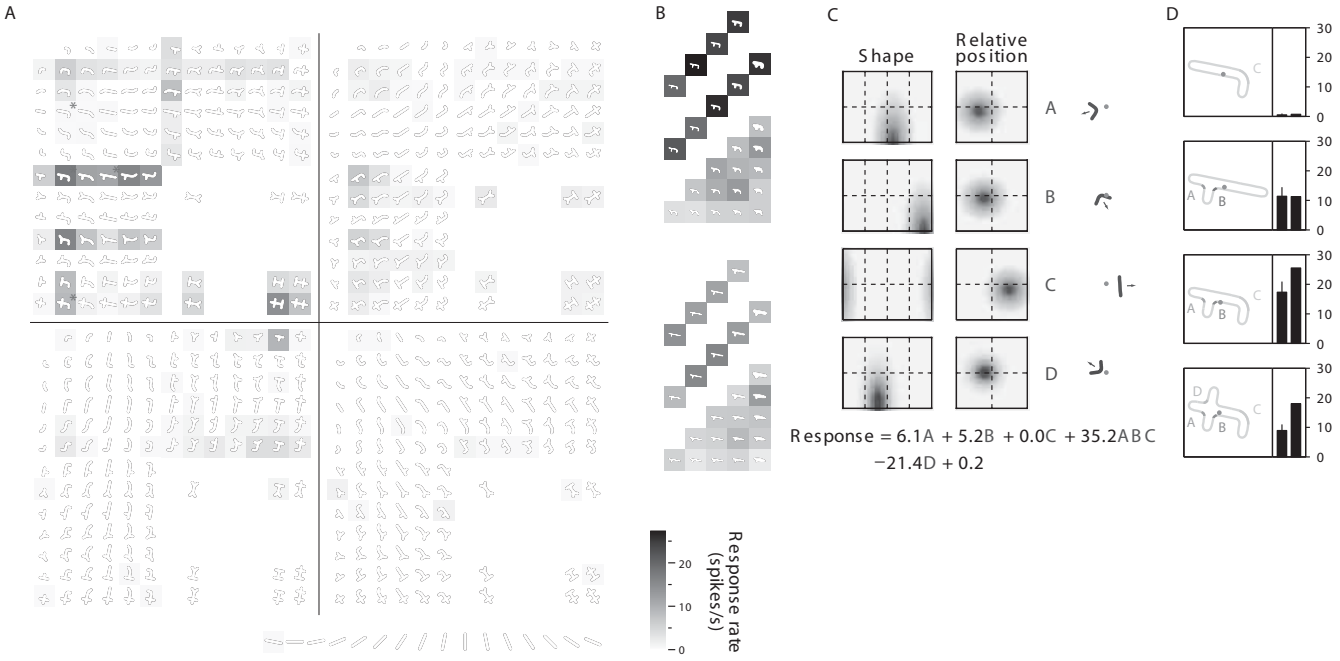


FIGURE 31.3 PIT neural tuning for boundary fragment configurations. (*A*) Average responses of a PIT neuron to stimuli constructed by factorial combination of boundary fragments. In this primary test, local curvature values were held constant. (*B*) Tuning for local curvature, tested with two representative shapes from the primary test. (*C*) Response model. The boundary fragment tuning dimensions for this model are orientation, curvature (mapped to a scale from −1.0 to 1.0), and *XY* position relative to stimulus center of mass. The best-fit model comprised three excitatory tuning regions (*A–C*) and one inhibitory tuning region (*D*). The equation at the bottom shows that the strongest response factor was the combined presence of all three excitatory boundary fragments. (*D*) Example stimuli showing the interactive effects of boundary fragments near the tuning peaks. In each case, the left bar in the histogram indicates observed response ± standard error, and the right bar indicates the response predicted by the model. (From Brincat & Connor (2004). Originally published in *Nature Neuroscience*.)

This kind of explicit single-neuron signal for configurations of disjoint parts is not envisioned in standard theories. One potential advantage of such signals is further compression of the object representation into a smaller number of more complex components. This would be particularly efficient if PIT neurons emphasize statistically common part configurations. At this scale, encompassing complete shapes, there is no simple geometric constraint on boundary structure, but common part configurations are bound to occur, owing to ecological factors. While this has not been investigated rigorously, it can be observed at an anecdotal level. The responses of the same example neuron to photographic stimuli (figure 31.4) illustrate how tuning for boundary fragment configurations might relate to common shape structures in natural object categories. In this case, the neuron responded to a leftward-facing quadruped (the polar bear), presumably owing to its combined sensitivity to opposed concavities on the left and broad curvature on the right. This is not to say that the neuron by itself signals the presence of quadrupeds. Rather, this neuron efficiently captures a common shape motif that would help to define quadrupeds as well as other object categories that have the same component structure. Other studies in more anterior parts of inferotemporal cortex suggest that responses to natural objects could be explainable in terms of structural components (Fujita, Tanaka, Ito, & Cheng, 1992; Perrett, Rolls, & Caan, 1982; Sigala & Logothetis, 2002; Tanaka, Saito, Fukada, & Moriya, 1991; Tsunoda, Yamane, Nishizaki, & Tanifuji, 2001; Wang, Fujita, & Murayama, 2000).

Single-neuron configuration signals might also enhance the cognitive accessibility of configural structure, supporting our ability to evaluate the physical potential of objects and interact with them in an accurate and intelligent manner. Finally, configural shape signals might also provide a basis for further integration, leading to global shape sensitivity at higher processing stages. This might be especially true for highly familiar or behaviorally relevant object categories that require maximally efficient processing, at the cost of dedicated neurons with extremely narrow selectivity. For generic object representation, representation in terms of component part configurations could be the optimal compromise between flexibility and efficiency.

One critical result from these PIT studies is the demonstration of spatial tuning in an object-centered reference frame. This is a central prediction of structural shape-coding theories (Biederman, 1987; Marr & Nishihara, 1978). Transformation from retinotopic to object-centered coordinates

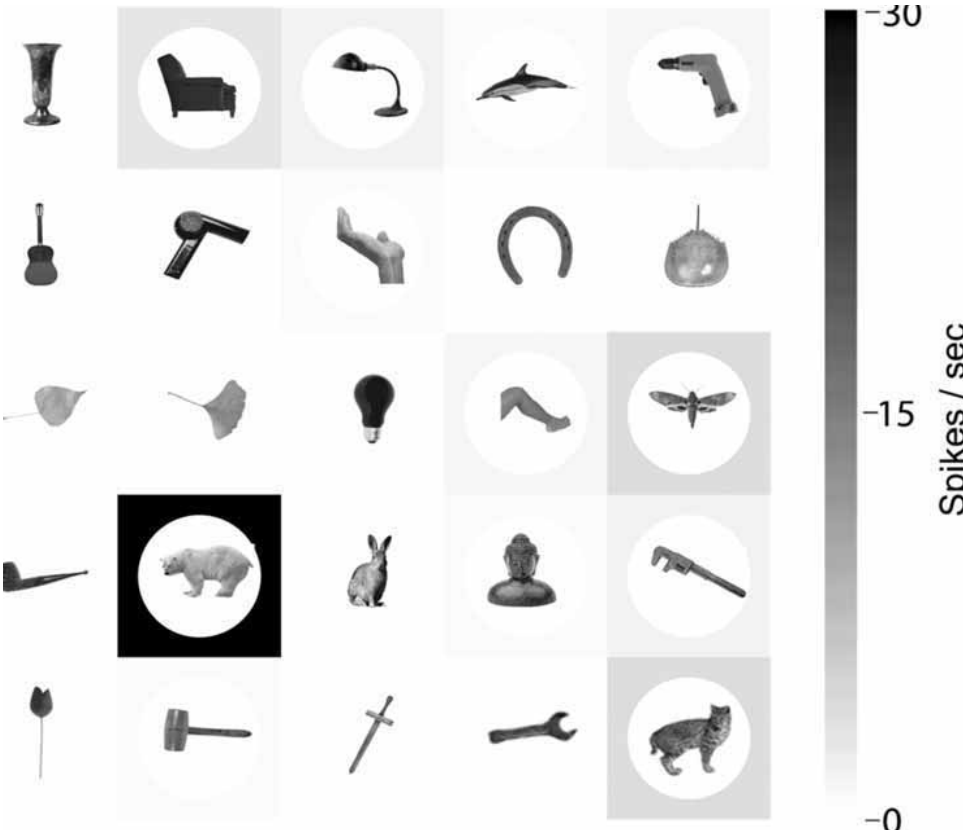


FIGURE 31.4 Responses of the same PIT neuron as in figure 31.3 to photographic stimuli. In each case, the average response to the photograph is indicated by the color of the background square (see scale bar at right).

achieves stability across changes in object position on the retina. Relative position signals in an object-centered reference frame provide the configural information that is required to distinguish different arrangements of the same parts. PIT neurons exhibit clear tuning for boundary fragment position with respect to object center. This is emphasized for the example neuron in figure 31.5, by showing the average response to stimuli containing the double-concavity configuration to the left of object center (top, strong responses) versus to the right (bottom, minimal responses). Figure 31.3*A* includes many stimuli with the same configuration in other object-centered positions (e.g., top or bottom) that likewise evoke little or no response. Control experiments (not shown) demonstrate that this tuning is consistent across retinotopic positions (within the PIT RF), so spatial tuning is much more acute in object-centered coordinates than in retinotopic coordinates.

Thus PIT neurons provide the kind of relative position information critical for parts-based structural coding. In addition to being centered on the object, the PIT reference frame probably scales with object size, given that PIT tuning functions (and tuning functions at higher stages in the ventral pathway; Ito, Tamura, Fujita, & Tanaka, 1995) exhibit remarkable consistency across size changes spanning multiple octaves. However, the spatial reference frame does not appear to rotate with the object, since shape tuning is not consistent across rotated versions of the same stimulus. This is exemplified in figure 31.3*A*, in which substantially rotated versions of the high-response shapes evoke no response whatsoever. Thus the theoretical prediction of a stable reference frame completely defined by the object is only partially confirmed at this level in the ventral pathway. Generalization across object rotation could be achieved in some other way, possibly by learning associations between different object views (Vetter, Hurlbert, & Poggio, 1995; Edelman & Poggio, 1991).

Three-dimensional object shape is represented in terms of surface fragments

The results described above relate to two-dimensional object boundary shape. Objects produce two-dimensional contrast boundaries in the retinal image; therefore it would be reasonable if two-dimensional boundary representation were the primary mode for object vision. In physical reality, however, objects are three-dimensional, and classical theories (Biederman, 1987; Marr & Nishihara, 1978) posit explicit representation of three-dimensional

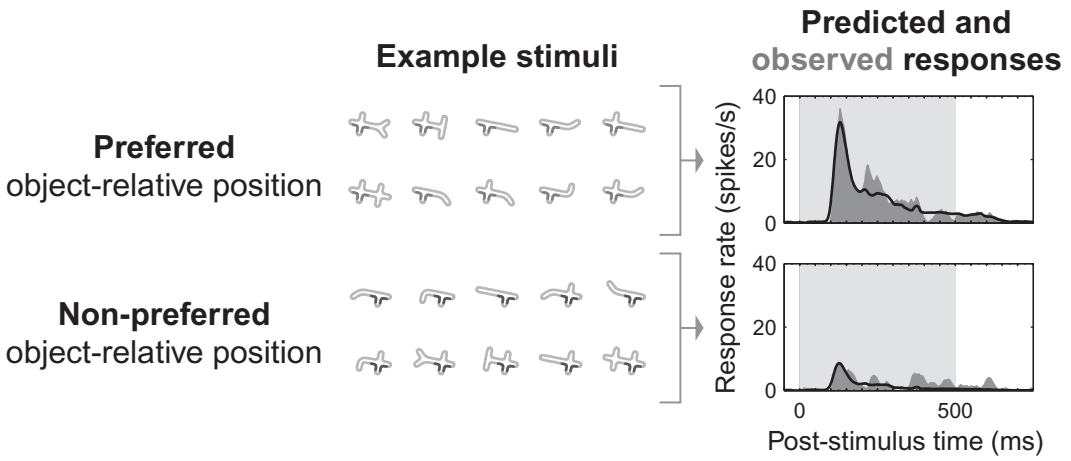


FIGURE 31.5 PIT tuning for object-centered position. Average responses of the same example neuron to two subsets of the stimuli in figure 31.3. In the top row, the opposed concavity configuration is positioned to the left of object center, and the average response is high, as shown by the peristimulus-time response histogram at the right. In the bottom row, the double concavity configuration is positioned to the right of object center, and the average response is low.

object structure, based on three-dimensional spatial configurations of three-dimensional parts. This is a particularly strong prediction, given the computational difficulty of extracting three-dimensional structure from two-dimensional retinal images and the higher-dimensional neural coding that is required to represent three-dimensional structure. In contrast, current computational vision models favor direct processing of two-dimensional images with no explicit representation of three-dimensional object structure (Fei-Fei, Fergus, & Perona, 2006; Lowe, 2004; Moghaddam & Pentland, 1997; Murase & Nayar, 1995; Riesenhuber & Poggio, 1999; Turk & Pentland, 1991; Weber, Welling, & Perona, 2000).

The three-dimensional structural coding hypothesis has not been directly tested, owing to the experimental difficulty of exploring the virtually infinite domain of three-dimensional object shape. In a recent attempt to overcome this obstacle, Yamane and colleagues (2008) used an evolutionary morphing strategy to sample three-dimensional object shape (figure 31.6). Neurons in central and anterior IT (CIT/AIT) were studied by first measuring their responses to an initial generation of 50 random three-dimensional shapes (figure 31.6, generation 1). The second stimulus generation included partially morphed descendants of higher-response stimuli from the first generation. This process was iterated across 8–10 generations, producing extensive sampling of stimuli in the high- and intermediate-response range of the cell. High-response stimuli were typically characterized by some shared local shape structure. In this case, the most noticeable shared structure is a ridge near the front of the shape facing out of the image plane.

When sampling was sufficiently complete, the response pattern could be used to constrain a quantitative model of the three-dimensional shape information encoded by the neuron. The best-fit model for this cell (figure 31.7) captured both the forward-facing ridge near the front of the object and the shallow concave dorsal surface behind it that characterized high-response stimuli. The response model was highly nonlinear; predicted (and observed) responses were substantial only for stimuli with both surface fragments. The result shown here typifies three-dimensional shape tuning observed for a substantial fraction of CIT/AIT cells. These neurons were tuned for three-dimensional spatial configurations of multiple surface fragments defined by their surface curvatures and three-dimensional orientations. These observations support the classic hypothesis that three-dimensional shapes are represented as structural configurations of three-dimensional parts.

In contrast to these findings regarding biological object vision, recent computational systems for object recognition have been most successful with nonstructural processing of two-dimensional image information (Fei-Fei et al., 2006; Lowe, 2004; Moghaddam & Pentland, 1997; Murase & Nayar, 1995; Riesenhuber & Poggio, 1999; Turk & Pentland, 1991; Weber et al., 2000). This makes sense, given the computational expense of inferring and encoding three-dimensional structure. It may be that even in the brain, rapid object recognition depends on two-dimensional processing (Hung, Kreiman, Poggio, & DiCarlo, 2005; Serre, Oliva, & Poggio, 2007) and neural coding of three-dimensional structure instead supports other aspects of object vision requiring detailed structural knowledge. Comparing similar objects within a recognized class, evaluating the functionality and utility of unfamiliar objects, anticipating physical events, and guiding physical interactions with objects are all likely to require detailed knowledge of three-dimensional structure.

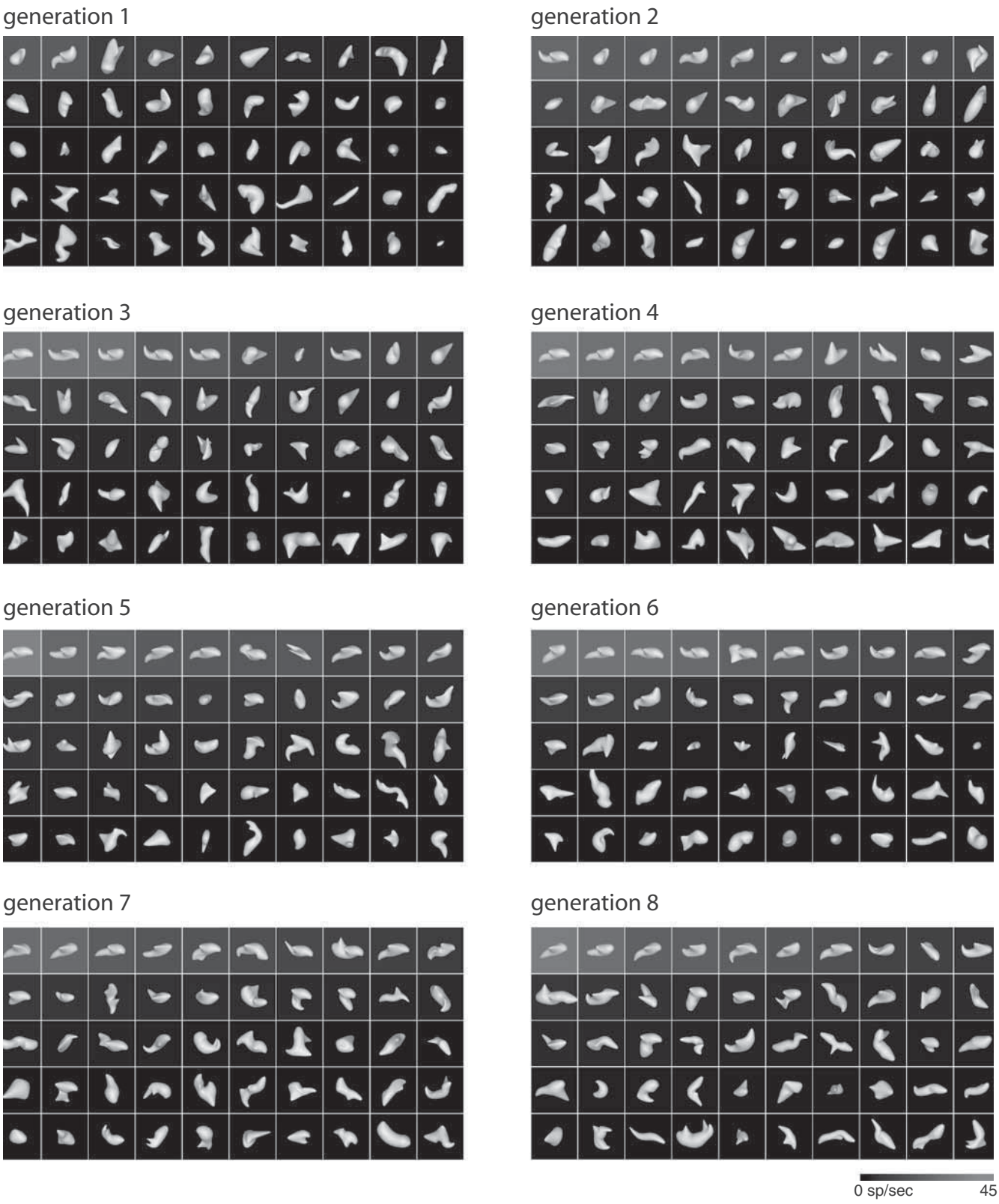

generation 1
generation 2
generation 3
generation 4
generation 5
generation 6
generation 7
generation 8
0 sp/sec
45

FIGURE 31.6 An evolutionary three-dimensional shape-morphing experiment. The first stimulus generation was created by randomly perturbing control points defining a topologically spherical spline mesh. Stimuli were rendered in three dimensions with shading cues (visible here) and corresponding stereoscopic (binocular disparity) cues. Stimuli were flashed at the center of gaze, and responses were recorded from a single neuron in AIT of a monkey performing a fixation task. For each stimulus, the background gray level denotes the neuron's average response (see the scale bar, lower right). Subsequent stimulus generations included partially morphed descendants of stimuli from previous generations. Ancestor stimuli were drawn in equal proportions from higher, intermediate, and lower response-level ranges. The end result was extensive sampling around the peak, shoulders, and boundaries of the neuron's shape-tuning range. (From Yamane et al. (2008). Originally published in *Nature Neuroscience*.)

Boundary fragment configurations are derived by recurrent network processing

The results described above suggest that the ventral pathway encodes objects as spatial configurations of boundary fragments. If so, this begs the more difficult question of how such configural information is derived. A direct answer to this question would require comprehensive measurement of neural network activity across multiple processing stages, which is not currently possible. But one indirect approach to inferring underlying network mechanisms is fine-scale temporal analysis of neural responses. If shape information is derived by time-consuming network processes, the evolution of that information across time (following stimulus onset) may be observable in the neural responses. As will be detailed below, the evolution of configural shape signals in PIT is observable in this way.

The figure 31.3 PIT neuron exemplifies supralinear integration of information across boundary fragments. As the equation in figure 31.3*C* reflects, responses produced by individual fragments were low (factors A, B, and C have coefficients below 7 spikes per second), while responses to fragment combinations were high (factor ABC has a large coefficient of 35 spikes per second). This nonlinear integration provides a selective, explicit signal for the overall configuration. In contrast, more linear summation of fragment information produces ambiguous signals associated with many different fragments or fragment combinations.

Fine-scale temporal analysis shows that PIT responses initially reflect linear summation, with nonlinear information emerging more gradually (Brincat & Connor, 2006). In many cases, this is observable in the responses of individual neurons. The PIT neuron represented in figure 31.8 was sensitive to concave boundary fragments oriented toward the upper left (135 degrees) and concave fragments oriented toward the lower right (315 degrees). The average temporal response profiles (solid gray histograms) for stimuli containing only the 315-degree concavity (top row) or only the 135-degree concavity (middle row) are phasic, confined to a window between 100 and 200 ms following stimulus onset. The response profile for stimuli containing both of these fragments (bottom row) includes an initial phasic spike in the 100- to 200-ms window that closely approximates the sum of the individual fragment responses (represented by the dark gray curve, which shows the predicted linear sum based on a temporal response model). For these stimuli, however, the response persists throughout the entire 500-ms stimulus

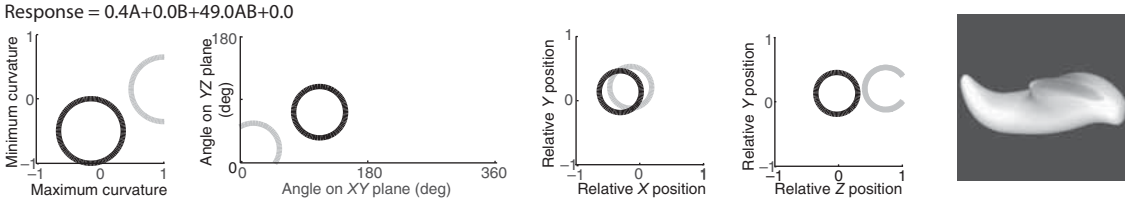


FIGURE 31.7 AIT neural tuning for three-dimensional surface fragment configuration. Each stimulus was defined in terms of its constituent surface fragments. Surface fragments were characterized in terms of their *XYZ* position relative to object center, three-dimensional orientation, and three-dimensional surface curvature (maximum and minimum cross-sectional curvature). Response patterns were fit with multiple Gaussian tuning functions in the position/orientation/curvature domain. The best-fit model for this cell was based on two Gaussian tuning regions (black and gray circles, describing 1.0 standard deviation boundaries). The gray tuning region defines surface fragments with sharp convex maximum curvature (near 1) and flat minimum curvature (near 0), that is, a ridge. The surface normal orientation points toward the viewer (near 0 on the *YZ*-plane). The position is toward the front of the object (near 1 on the *Z*-axis). The surface region on a high response stimulus (at right) corresponding to this Gaussian function is tinted gray. The other Gaussian tuning region (black) defines shallow concave surfaces with normals pointing upward (near 90 degrees on the *XY*- and *YZ*-planes) positioned near object center. The response equation at the top indicates low responses to stimuli with only one surface fragment or the other (A or B) but high responses for the combination (AB). (From Yamane et al. (2008). Originally published in *Nature Neuroscience*.)

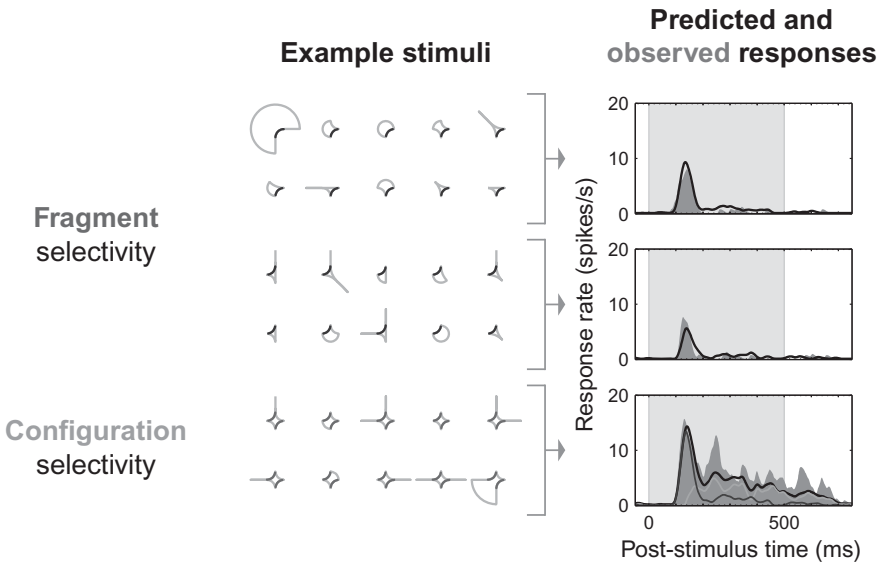


FIGURE 31.8 Time course of linear and nonlinear boundary fragment responses for a PIT neuron. *Top row*, Average response to stimuli containing a concavity oriented toward 315 degrees. The time course of observed response (gray histogram) and response predicted by a temporal model (black curve) are shown. The gray box indicates the stimulus presentation period. *Middle row*, Average response to stimuli containing a concavity oriented toward 135 degrees. *Bottom row*, Average response to stimuli containing both concavities. The total predicted response (black curve), predicted linear response due to individual fragment terms (dark gray curve), and predicted nonlinear response due to the fragment combination term (light gray curve) are shown. (Adapted from *Neuron* (Brincat & Connor, 2006) with permission from Elsevier.)

presentation. Thus beyond 200 ms following stimulus onset, this neuron exhibits highly nonlinear integration and conveys an explicit signal for the necklike configuration of two opposed concavities.

This example reflects a general trend in PIT for linear boundary fragment responses to evolve more quickly, peaking around 120 ms after stimulus onset (figure 31.9*A*; the black curve summarizes linear response strength across a sample of 89 PIT neurons). In contrast, nonlinear response strength evolved more gradually, peaking 180 ms after onset (figure 31.9*A*, gray curve). These trends were partly due to single-neuron tuning transitions as in figure 31.8 (figure 31.9*B*, thin curves) and partly due to differential response profiles of consistently linear neurons (figure 31.9*B*, thick black curve) versus consistently nonlinear neurons (figure 31.9*B*, thick gray curve). This overall pattern is consistent with a fairly simple neural network model in which neurons vary in the relative strength of V4-like boundary fragment inputs and recurrent inputs (recurrent excitatory inputs from cells with similar configuration tuning, recurrent inhibitory inputs from cells with dissimilar tuning) (Brincat & Connor, 2006; Salinas & Abbott, 1996). Neurons with stronger V4 inputs respond quickly and in a more linear fashion, while neurons with stronger recurrent connectivity respond more slowly and in a more nonlinear fashion. The 60-ms delay for part configuration signals in PIT is consistent with a remarkably similar delay for pattern motion signals in area MT (Movshon, Adelson, Gizzi, & Newsome, 1985; Rodman & Albright, 1989; Pack, Berezovskii, & Born, 2001; Smith, Majaj, & Movshon, 2005). In both cases, it could be that some kind of recurrent network processing is required to generate unambiguous signals based on integration across multiple stimulus components. There are, however, alternative interpretations that cannot be ruled out at this point. For example, selectivity for combined inputs might be produced by a static threshold nonlinearity and could be temporarily masked by transient onset responses.

Summary: Configural representation of object structure

The findings reviewed above suggest at least a partial explanation of how the ventral visual pathway achieves compact, stable representation of useful object information. In area V4, object boundary fragments are summarized in terms of their first- and second-order derivatives (orientation and curvature), greatly reducing the dimensionality of the retinal response pattern. At the next processing stage in PIT, configurations of multiple fragments are represented (further reducing dimensionality) in an object-centered reference frame (producing stability across changes in retinal position). Recent results in more anterior regions of inferotemporal cortex suggest that this configural coding scheme generalizes to three-dimensional shape representation in terms of object surface fragments. Considering the processing time that is required to perfect these configural representations (on the

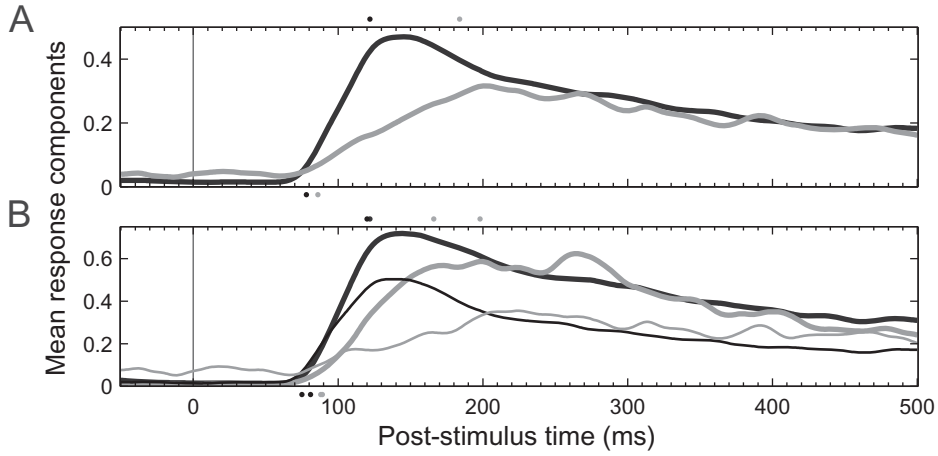


FIGURE 31.9 The average time course of linear and nonlinear response components in PIT. (*A*) Normalized linear (black curve) and nonlinear (gray curve) response strength averaged across temporal models fit to responses of 89 PIT neurons. Dots with corresponding colors indicate the estimated onset and peak (90% maximum) times for linear and nonlinear responses. (*B*) The same linear and nonlinear response strength averages are partitioned into neurons with consistently linear responses across time (thick black curve), neurons with consistently nonlinear responses (thick gray curve), and linear (thin black curve) and nonlinear (thin gray curve) model components for neurons that transitioned across time. (Adapted from *Neuron* (Brincat & Connor, 2006) with permission from Elsevier.)

order of 200 ms), they might not explain the most rapid human recognition speeds (Thorpe, Fize, & Marlot, 1996). However, these rapid reaction times could be limited to coarse categorization (Rousselet, Mace, & Fabre-Thorpe, 2003), based on nonconfigural parts-level information available after about 100 ms (figure 31.9*A*, black curve). Finer discrimination based on larger-scale shape configurations rather than diagnostic parts requires longer processing times (Arguin & Saumier, 2000; Wolfe & Bennett, 1997; Ringach & Shapley, 1996), consistent with the delayed emergence of explicit configural signals (figure 31.9*A*, gray curve). Aside from recognition and discrimination, configural representations could support other important aspects of object vision, including cognitive evaluation of object structure and function as well as guidance of precise physical interactions with complex objects.

REFERENCES

ARGUIN, M., & SAUMIER, D. (2000). Conjunction and linear nonseparability effects in visual shape encoding. *Vis. Res.*, *40*(22), 3099–3115.

BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychol. Rev.*, *94*(2), 115–147.

BRINCAT, S. L., & CONNOR, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.*, *7*(8), 880–886.

BRINCAT, S. L., & CONNOR, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron*, *49*(1), 17–24.

EDELMAN, S., & POGGIO, T. (1991). Models of object recognition. *Curr. Opini. Neurobiol.*, *1*(2), 270–273.

FEI-FEI, L., FERGUS, R., & PERONA, P. (2006). One-shot learning of object categories. *IEEE Trans. Pattern Analysis Machine Intelligence*, *28*(4), 594–611.

FELLEMAN, D. J., & VAN ESSEN, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, *1*(1), 1–47.

FUJITA, I., TANAKA, K., ITO, M., & CHENG, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, *360*(6402), 343–346.

GATTAS, R., SOUSA, A. P. B., & GROSS, C. G. (1988). Visuotropic organization and extent of V3 and V4 of the macaque. *J. Neurosci.*, *8*(6), 1831–1845.

HUBEL, D. H., & WIESEL, T. N. (1959). RFs of single neurones in the cat's striate cortex. *J. Physiol.*, *148*, 574–591.

HUBEL, D. H., & WIESEL, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.*, *28*(2), 229–289.

HUBEL, D. H., & WIESEL, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.*, *195*(1), 215–243.

HUNG, C. P., KREIMAN, G., POGGIO, T., & DICARLO, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, *310*(5749), 863–866.

ITO, M., TAMURA, H., FUJITA, I., & TANAKA, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.*, *73*(1), 218–226.

LOWE, D. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, *60*(2), 91–110.

MARR, D., & NISHIHARA, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. B Biol. Sci.*, *200*(1140), 269–294.

MILNER, P. M. (1974). A model for visual shape recognition. *Psychol. Rev.*, *81*(6), 521–535.

MOGHADDAM, B., & PENTLAND, A. (1997). Probabilistic visual learning for object representation. *IEEE Trans. Pattern Analysis Machine Intelligence*, *19*(7), 696–710.

MOVSHON, J. A., ADELSON, E. H., GIZZI, M. S., & NEWSOME, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern recognition mechanisms* (pp. 117–151). New York: Springer.

Murase, H., & Nayar, S. K. (1995). Visual learning and recognition of 3-D objects from appearance. *Int. J. Comput. Vis.*, *14*, 5–24.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Pack, C. C., Berezovskii, V. K., & Born, R. T. (2001). Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature*, *414*(6866), 905–908.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *J. Neurophysiol.*, *82*(5), 2490–2502.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *J. Neurophysiol.*, *86*(5), 2505–2519.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nat. Neurosci.*, *5*(12), 1332–1338.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.*, *47*(3), 329–342.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.*, *2*(11), 1019–1025.

Ringach, D. L., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vis. Res.*, *36*(19), 3037–3050.

Rodman, H. R., & Albright, T. D. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Exp. Brain Res.*, *75*(1), 53–64.

Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *J. Vis.*, *3*(6), 440–455.

Salinas, E., & Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex. *Pro. Natl. Acad. Sci. USA*, *93*(21), 11956–11961.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *The mechanization of thought process*. London: H.M. Stationery Office.

Serre, T., Oliva, A., & Poggio, T. (2007). A feed forward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. USA, 104*(15), 6424–6429.

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–320.

Smith, M. A., Majaj, N. J., & Movshon, J. A. (2005). Dynamics of motion signaling by neurons in macaque area MT. *Nat. Neurosci.*, *8*(2), 220–228.

Sutherland, N. S. (1968). Outlines of a theory of visual pattern recognition in animals and man. *Proc. R. Soc. Lond. B Biol. Sci.*, *171*, 297–317.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.*, *66*(1), 170–189.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.

Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat. Neurosci.*, *4*(8), 832–838.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *J. Cogn. Neurosci.*, *3*(1), 71–86.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. G. Ingle, M. A. Goodale, & R. J. Q. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Vetter, T., Hurlbert, A., & Poggio, T. (1995). View-based models of 3D object recognition: Invariance to imaging transformations. *Cereb. Cortex*, *5*(3), 261–269.

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*(5456), 1273–1276.

Wang, Y., Fujita, I., & Murayama, Y. (2000). Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. *Nat. Neurosci.*, *3*(8), 807–813.

Weber, M., Welling, M., & Perona, P. (2000). Towards automatic discovery of object categories. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (Vol. 2, pp. 101–108).

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vis. Res.*, *37*(1), 25–43.

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., & Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.*, *11*(11), 1243–1244.

32 The Cognitive and Neural Development of Face Recognition in Humans

ELINOR McKONE, KATE CROOKES, AND NANCY KANWISHER

ABSTRACT Conventional wisdom has long held that face recognition develops very slowly throughout infancy, childhood, and adolescence, with perceptual experience as the primary engine of this development. However, striking new findings from just the last few years have overturned much of this traditional view by demonstrating genetic influences on the face recognition system as well as impressive face discrimination abilities that are present in newborns and in monkeys that were reared without ever seeing a face. Nevertheless, experience does play a role, for example, in narrowing the range of facial subtypes for which discrimination is possible and perhaps in increasing discrimination abilities within that range. Here we first describe the cognitive and neural characteristics of the adult system for face recognition, and then we chart the development of this system over infancy and childhood. This review identifies a fascinating new puzzle to be targeted in future research: All qualitative aspects of adult face recognition measured behaviorally are present very early in development (by 4 years of age; all that have been tested are also present in infancy), yet functional magnetic resonance imaging and event-related potential evidence shows very late maturity of face-selective neural responses (with the fusiform face area increasing substantially in volume between age 7 years and adulthood).

Introduction

One of the most impressive skills of the human visual system is our ability to identify a specific individual from a brief glance at their face, thus distinguishing that individual from hundreds of other people we know, despite the wide variations in the appearance of each face as it changes in viewpoint, lighting, emotional expression, and hairstyle. Though many mysteries remain, important insights have been gleaned over the last two decades about the cognitive and neural mechanisms that enable humans to recognize faces. Here, we address an even more difficult and fundamental question: How does the machinery of face recognition get wired up during development in the first place?

Our review of the available evidence supports a view of the development of face recognition that is dramatically different from the one suggested by the first studies in the field. Twenty years ago, the standard theory was that core aspects of the ability to discriminate faces were not present until 10 years of age, and their emergence and eventual maturity were determined primarily by experience (Carey & Diamond, 1977; Carey, Diamond, & Woods, 1980). This position has been overturned by recent findings that demonstrate striking abilities even in neonates and by mounting evidence of genetic contributions.

We organize our review by age group. Throughout, we ask how the available data address the following fundamental theoretical questions:

1. What are the inherited genetic contributions to the specification of the adult system for processing facial identity information?
2. What is derived from experience?
3. How exactly do genes and/or experience work separately or together across the course of development to produce the adult system?

The perception of face identity in adulthood

We begin with a characterization of the end state of development: the cognitive and neural basis of the perception of facial identity in adults. Note that this is a major topic in its own right, with much internal theoretical debate. However, to facilitate our present interest in the *developmental* course of face recognition, we focus on empirical phenomena, especially those that are well established in adults and have subsequently been tested in development.

CORE BEHAVIORAL PROPERTIES OF FACE IDENTITY PERCEPTION IN ADULT HUMANS Basic properties of face identification in adults are as follows. Identification is more accurate when faces are upright than when they are inverted (i.e., upside

ELINOR McKONE and KATE CROOKES Department of Psychology, Australian National University, Canberra, Australia
NANCY KANWISHER McGovern Institute for Brain Research and Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, Massachusetts

down) on both memory and perceptual tasks, and the inversion decrement is substantially larger for faces than for nonface objects (the *disproportionate inversion effect*; Yin, 1969; see also Robbins & McKone, 2007). Generalization from a single image of a novel face in one viewpoint to an image in another is relatively poor, albeit better from the three-quarter view to front or profile views than between the more distinct profile and front views (the *three-quarter view advantage*; Logie, Baddeley, & Woodhead, 1987). For familiar faces, performance on memory tasks relies more strongly on inner face regions than on external regions that include hair; for unfamiliar faces, the pattern is reversed (*inner versus outer features effects*; Ellis, Sheperd, & Davies, 1979). Finally, identification of own-race faces is better than identification of other-race faces (the *other-race effect*; Meissner & Brigham, 2001). Note that the first two properties (i.e., the disproportionate inversion effect and the three-quarter view advantage) derive directly from perceptual processing, but the last two are known to derive at least partly from deliberate task strategies (e.g., reliance on hair for novel faces if distinctive hair is present; Duchaine & Weidenfeld, 2003) or social and attentional factors (other-race effect; Bernstein, Young, & Hugenberg, 2007).

Additional experimental findings can be grouped under the heading of phenomena that have motivated the concept of *holistic/configural* processing. Holistic/configural processing is defined (e.g., Tanaka & Farah, 1993; Maurer, Lewis, & Mondloch, 2005) as (1) a strong integration at the perceptual level of information from all regions of the face (so that altering one region leads to changes in the percept of other regions), which (2) codes the exact spacing between face features (and, more controversially, exact feature shape as well; Yovel & Duchaine, 2006), and (3) is strongly sensitive to face inversion. Relevant phenomena are as follows. Subjects find it harder to identify one half of a combination face (e.g., the top half of George Bush's face with the bottom half of Tony Blair's face) if the inconsistent other half-face is aligned with the target half rather than misaligned (the *composite effect*; Young, Hellawell, & Hay, 1987). Subjects are also better able to distinguish which of two face parts (e.g., two noses) appeared in a previously shown face when these are presented in the context of the whole face than when they are presented in isolation (the *part-whole effect*; Tanaka & Farah, 1993). Part choice is also better in the original whole than in a version of the whole face with an alteration in spacing between nontarget features (the *part-in-spacing-altered-whole effect*; Tanaka & Sengco, 1997), a finding that is consistent with other evidence of strong *sensitivity to spacing changes* (e.g., distance between eyes) in upright faces (e.g., Rhodes, Brake, & Atkinson, 1993; McKone, Aitkin, & Edwards, 2005). When an upright and an inverted version of a face are superimposed in transparency, the upright face is perceived more strongly (*perceptual bias to upright*; Martini, McKone, & Nakayama, 2006). All these holistic effects are specific to upright faces; they are not found for inverted or scrambled faces (Young et al., 1987; Tanaka & Sengco, 1997; Robbins & McKone, 2003; Martini et al., 2006) and are weak or absent for objects, including objects of expertise (for reviews, see McKone, Kanwisher, & Duchaine, 2007; Robbins & McKone, 2007).

Finally, other behavioral phenomena have been taken to indicate coding within a perceptual "face-space," defined as a multidimensional space in which each individual face is coded as a point by its value on underlying dimensions describing different aspects of facial structure and for which the "average" face lies at the center of the space (Valentine, 1991). These phenomena include *distinctiveness effects*, in which performance is better for distinctive faces than for typical faces on old-new recognition tasks but the pattern is reversed on face versus nonface classification tasks (Valentine & Bruce, 1986), and *adaptation aftereffects*, in which, for example, adaptation to expanded faces make a physically normal face appear contracted (Webster & MacLin, 1999) and adaptation to "anti-Bill" (the physical opposite of Bill in face space) makes the average face appear like Bill (Leopold, O'Toole, Vetter, & Blanz, 2001).

Neurophysiology and Functional Magnetic Resonance Imaging in Adult Monkeys Adult monkeys show cortical mechanisms specialized for face perception. Strongly face-selective responses from single neurons ("face cells") are well established in the temporal lobes of macaques (Desimone, Albright, Gross, & Bruce, 1984; Foldiak, Xiao, Keysers, Edwards, & Perrett, 2004), and face-selective cortical regions have been reported in macaques using functional magnetic resonance imaging (fMRI) (Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Pinsk, DeSimone, Moore, Gross, & Kastner, 2005). Tsao, Freiwald, Tootell, and Livingstone (2006) demonstrated direct correspondence between face-selective fMRI patches and face selectivity of single cells within those patches. Note that the role of "face cells" in supporting the behavioral phenomena described in the previous section is mostly unexplored, with the exceptions that a preponderance of face-selective cells are tuned to upright (Perrett et al., 1988) and that their tuning to facial distortions from the "average face" is consistent with a face space coding of facial identity (Leopold, Bondar, & Giese, 2006). In development, only basic face selectivity has been studied.

Functional Magnetic Resonance Imaging: Cortical Loci of Face Identity Processing in Adult Humans Brain imaging in humans reveals three face-selective cortical regions (figure 32.1), of which the fusiform face area (FFA) (Kanwisher, McDermott, & Chun, 1997), is the main one that is investigated in children. This region, which can be

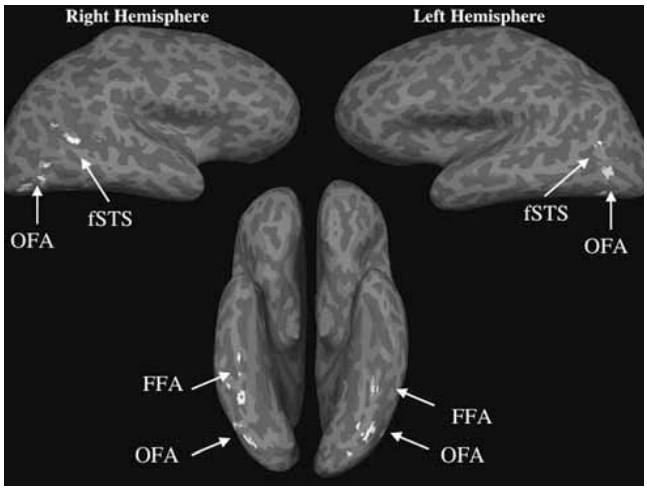


FIGURE 32.1 Face-selective activation (faces > objects, $p < 0.0001$) on an inflated brain of one adult subject, shown from lateral and ventral views of the right and left hemispheres. Three face-selective regions are shown: the FFA in the fusiform gyrus along the ventral part of the brain, the OFA in the lateral occipital area, and the fSTS in the posterior region of the superior temporal sulcus. For studies of face identification (rather than expression, etc.), the FFA and OFA are of greatest interest. (See color plate 46.)

found in essentially every normal adult in a short "localizer" scan (Saxe, Brett, & Kanwisher, 2006), responds more strongly to faces than to letter strings and textures (Puce, Allison, Asgari, Gore, & McCarthy, 1996), flowers (McCarthy, Luby, Gore, & Goldman-Rakic, 1997), and indeed all other nonface stimuli that have been tested to date, including mixed everyday objects, houses, hands (Kanwisher et al., 1997), and objects of expertise (Kanwisher & Yovel, in press).

fMRI adaptation studies show that neural populations in the FFA can discriminate face identity (Rotshtein, Henson, Treves, Driver, & Dolan, 2005) but not facial expression (Winston, Vuilleumier, & Dolan, 2003). The FFA is involved in individual discrimination of upright but not inverted faces (Yovel & Kanwisher, 2005; Mazard, Schiltz, & Rossion, 2006), and its inversion effect (i.e., higher response to upright than inverted faces) correlates with the behavioral inversion effect (Yovel & Kanwisher, 2005). The FFA also demonstrates holistic processing, specifically a composite effect (Schiltz & Rossion, 2006).

ELECTROPHYSIOLOGICAL SIGNATURES IN HUMAN ADULTS A negative-going event-related potential (ERP) response peaking about 170 ms after stimulus onset over posterior temporal sites (N170) has been widely replicated to be face-selective (Halgren, Raij, Marinkovic, Jousmaki, & Hari, 2000; Liu, Harris, & Kanwisher, 2002). This peak is delayed by 10 ms, and is larger in amplitude for inverted faces than for upright faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996). The N170 also shows identity discrimination (a lower response for repeated compared to unrepeated faces) when the faces are upright but not inverted (Jacques & Rossion, 2006; Jacques, D'arripe, & Rossion, 2007). An important point that is relevant to the interpretation of developmental studies is that the neural source of the N170 is unknown even in adults, and the sources of suggested equivalent components in children and infants could possibly be different still.

Data from adult subjects relevant to the roles of experience and genetics

Before considering what developmental studies tell us about the roles of experience and genetics in face recognition, we describe several findings from adults that also bear directly upon these issues.

Clearly, *experience in isolation* can influence face perception. Adults continue to learn new faces throughout life, and this improves perceptual discrimination of these faces: Matching the correct face photograph to a degraded security camera video image is more accurate if the face is familiar than if it is unfamiliar (Burton, Wilson, Cowan, & Bruce, 1999; also see Bruce, Henderson, Newson & Burton, 2001). Temporary aftereffects from adaptation to distorted faces (e.g., Webster & MacLin, 1999) also indicate purely experience-based changes in the tuning of perceptual representations of faces. Training effects on the ability to discriminate trained and novel faces have also been demonstrated in an adult prosopagnosic (DeGutis, Bentin, Robertson, & D'Esposito, 2007). Interestingly, however, there is no evidence that experience alone produces any fundamental *qualitative* change in face processing either neurally or cognitively; for example, holistic processing, "face space" effects, and FFA activation all occur strongly for both familiar faces and unfamiliar faces (Young et al., 1987; Kanwisher et al., 1997; Webster & MacLin, 1999; Le Grand, Mondloch, Maurer, & Brent, 2004; Carbon et al., 2007).

Studies of human adults provide two sources of evidence for genetic contributions. Inability to recognize faces in the absence of any known brain injury (developmental prosopagnosia) often runs in families (Duchaine, Germine, & Nakayama, 2007; Grueter et al., 2007; Kennerknecht, Pluempe, & Welling, 2008). Also, in normal adults, fMRI shows greater similarity in the pattern of activation across the ventral visual stream for monozygotic compared to dizygotic twins, but only for stimulus classes for which an evolutionary origin of the observed selective cortical regions could reasonably be proposed: faces and places but not written words or chairs (Polk, Park, Smith, & Park, 2007).

In summary, results from adults tell us that experience can fine-tune face recognition without changing its qualitative properties and that genes explain some of the variation behaviorally and neurally. Importantly, adult studies do not tell us at what developmental stage genes have their

influence. In particular, they do not necessarily demonstrate that a face system is present at birth. Some genetically predetermined processes are present at birth (e.g., the sucking reflex), but others affect maturational processes later in childhood or adolescence (e.g., puberty).

Development: Infancy

In exploring genetic and experience-based contributions to face recognition via infancy studies, several interrelated questions are relevant. First, which abilities, if any, are present at birth? Visual abilities that are present in neonates (or in monkeys that have been deprived of all face input) cannot be derived from experience and therefore provide the only method of revealing genetic influences in isolation from any visual learning. Second, if babies are born with a face representation, is its purpose merely to draw attention to faces (cf. CONSPEC in Morton & Johnson, 1991) or to support individuation? Third, how broadly tuned is any such representation: broad enough to cover any primate face, specific to own-species faces, or perhaps even to own-race faces? Finally, which, if any, of the types of effects of experience in early infancy that are found in other perceptual and cognitive domains occur for faces: Improvements with increasing experience? Perceptual narrowing (i.e., destruction of earlier ability)? Critical periods? Studies of these topics published within the last few years have dramatically altered our understanding of infant face recognition.

In a classic result, newborns (median age: 9 minutes) track an upright "paddle face" (figure 32.2*A*) further than versions in which the position of the internal blobs is scrambled or inverted (Goren, Sarty, & Wu, 1975; Johnson, Dziurawiec, Ellis, & Morton, 1991). Although it has been suggested that this preference could arise from general visual biases (e.g., for stimuli with more elements in the upper visual field; Simion, Macchi Cassia, Turati, & Valenza, 2003), preference only for the normal contrast polarity of a (Caucasian) face (Farroni et al., 2005) argues for a level of specificity to facelike structure. Thus humans are born with some type of innate preference that, at the very least, attracts infants' attention to faces. Note that the innate representation supporting face preference could be different from that supporting face individuation in adults (Johnson, 2005); indeed, a finding that neonates track faces in the temporal but not nasal visual field (Simion, Valenza, Umlita, & Dalla Barba, 1998) suggests a subcortical rather than cortical origin.

Our concern in this chapter is primarily with the development of face individuation ability. This can be measured in infants by looking time measures that assess preference and dishabituation-to-perceived-novelty. A classic finding is that neonates less than 4 days old can discriminate their mother from similar-looking women (Pascalis, de Schonen, Morton, Deruelle, & Fabre-Grenet, 1995; Bushnell, 2001), although

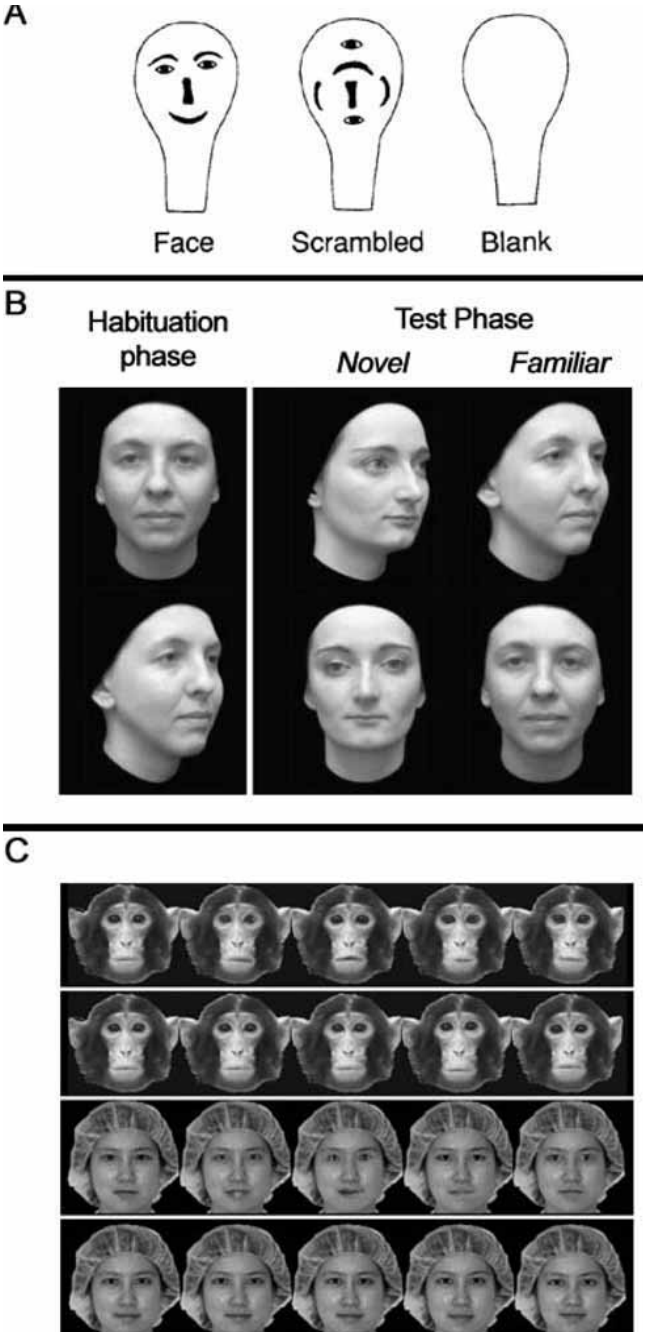


FIGURE 32.2 Face perception without experience. (*A*) Newborn humans (<1 hour old) track the "paddle face" on the left further than the scrambled version (Morton & Johnson, 1991). (*B*) Newborn humans (<3 days) look longer at the novel than habituated face, indicating recognition of face identity even across view change (Turati et al., 2008). (*C*) Japanese macaques raised with no exposure to faces can, on first testing, discriminate very subtle differences between individual monkey faces (including differences both in shape and in spacing of internal features) and can also do this for human faces (Sugita, 2008).

mother recognition in the first 24 hours may be partially dependent on prenatal familiarity with her voice (Sai, 2005). More recent data demonstrate even more striking abilities. Three-month-olds can recognize the identity of *novel* individuals, with *similar-looking* faces (same sex, age, race), *without hair*, and *across view changes* (Pascalis, De Haan, Nelson, & de Schonen, 1998; Kelly et al., 2007). Indeed, it has very recently been discovered that newborns (<3 days) can perform this task (Turati, Bulf, & Simion, 2008; see figure 32.2*B*). Moreover, the newborns discriminated only front to three-quarter view changes and not three-quarter to profile, in a pattern somewhat (although not precisely) similar to the three-quarter view advantage that is seen in adults. Finally, newborns demonstrate an inversion effect on discrimination, with babies 1–3 days old discriminating same-view faces without hair upright but not inverted (Turati, Macchi Cassia, Simion, & Leo, 2006).

The newborn discrimination findings strongly suggest that a face representation, tuned to upright and able to support individual-level representation, is present at birth. It seems unlikely that 3 "days" of experience with faces—in fact, a maximum of perhaps 12 hours of visual experience of any kind (newborns sleep 16 hours per day and have their eyes shut during breastfeeding and crying)—would be sufficient for a purely learning-based system to support the level of fine discrimination ability that is observed.

Even more compelling, however, is a recent behavioral study in monkeys (Sugita, 2008). Japanese macaques were raised by human caregivers wearing masks, giving the monkeys no exposure to faces but otherwise normal visual experience in a complex environment. On their *very first* experience with faces (aged 6–24 months), the monkeys showed a preference to look at static photographs of faces over photographs of objects that were equally novel in their visual environment (e.g., cars, houses) and discriminated very subtle differences between individual faces (figure 32.2*C*) in a habituation paradigm.

A variety of other infant findings also either directly argue that a representational capacity for differentiating individual face structures is present at birth or at least do not reject this conclusion. Newborns (<1 week) prefer faces rated by adults as attractive over unattractive faces when the faces are upright but not inverted (Slater, Quinn, Hayes, & Brown, 2000). Regarding holistic processing, Sugita's (2008) monkeys discriminated spacing changes (figure 32.2*C*) with almost no prior experience of faces (they had been exposed to faces only during the short face preference task), and five-month-old humans discriminate spacing changes small enough to fall within the normal physical range, upright but not inverted (Hayden, Bhatt, Reed, Corbly, & Joseph, 2007); also babies 6–8 months old show a composite-like effect in which the combination of the inner features of one old face with the outer features of another old face is treated as a new individual, upright but not inverted (Cohen & Cashon, 2001). At 3 months (although not at 1 month), human infants falsely recognize the average of four studied faces as "old," a phenomenon that is also shown by adults (de Haan, Johnson, Maurer, & Perrett, 2001). Importantly, there are no major behavioral properties of face recognition present in adults that are known *not* to be present in infants; where we have not mentioned properties (e.g., adaptation aftereffects), this is because no relevant data exist, not because infants have been tested and failed to show effects.

Findings of *perceptual narrowing* indicate that (1) a representational capacity for faces that is present at birth can initially be applied to a wide range of faces but that (2) this range gets restricted during the first several months of life to include only the kinds of faces (i.e., species or race) that have been seen in this period. Perceptual narrowing is best known from the domain of language (e.g., Kuhl, Tsao, & Liu, 2003). Infants are born with the ability to discriminate phoneme boundaries from all possible languages in the world (e.g., English and Japanese), but over the first 6–12 months of life, they lose the ability to discriminate phonemes from nonexperienced languages (e.g., Japanese for a child from a monolingual English-speaking family), and even extensive exposure as an adult is usually insufficient to regain native-speaker levels of discrimination and reproduction. For faces, five studies have reported and explored properties of perceptual narrowing. In humans, Pascalis, de Haan, and Nelson (2002) showed that 6-month-old infants could discriminate both human and monkey faces, while 9-month-olds and adults could discriminate only human faces. Kelly and colleagues (2007) reported that Caucasian babies from the north of England, with high exposure to Caucasians but essentially no exposure to African or Asian faces, could recognize individuals (across view change) from all three races at 3 months of age. At 6 months, Caucasian babies could no longer individuate African faces; at 9 months, they had additionally lost the ability to individuate Asians. The Sugita (2008) study described earlier reported that on first exposure to faces, the monkeys not only could discriminate individual monkey faces (other macaques), but also could make extremely fine discriminations among human faces (figure 32.2*C*). Following 1 month of exposure to a single face type (either human or monkey, involving live interaction for least 2 hours per day), Sugita's monkeys lost the ability to discriminate individuals of the nonexperienced species. Relearning was also difficult; monkeys that were initially exposed only to humans failed to discriminate monkey faces even after subsequently sharing a cage with 10 other monkeys for 11 months. (Note, however, that there is some evidence of flexibility in humans into middle childhood: Korean children adopted to Caucasian Francophone countries at age 3–9 years showed, as adults, better recognition memory for Caucasian faces than for Korean faces; Sangrigoli, Pallier,

Argenti, Ventureyra, & de Schonen, 2005). During human infancy, perceptual narrowing can be avoided by deliberate exposure to face types that the infant would not naturally see; regular exposure to monkey faces beginning at 6 months leads to retained ability to discriminate monkey faces at 9 months (Pascalis et al., 2005). Perceptual narrowing for faces also has an interesting possible link with narrowing for language. Lewkowicz and Ghazanfar (2006) reported that human infants could make cross-modality matches of a monkey vocalization to a picture of a monkey face making that particular sound at 4 and 6 months but that this ability was lost at 8 and 10 months.

Importantly, the perceptual narrowing effects for faces described above indicate only a destructive effect of experience across infancy (i.e., loss of initial ability with other species and other races). In the domain of language, loss of phonetic discrimination ability within nonexperienced languages has been shown to co-occur with an improvement of phonetic discriminability within the experienced language (Kuhl et al., 2006). Thus perceptual narrowing for faces might similarly include enhanced ability to discriminate experienced face subtypes; that is, discrimination for own-species own-race faces might start crude and *improve* with practice. Potentially consistent with this prediction, Humphreys and Johnson (2007) showed that the physical difference between faces that was required to produce novelty preference was smaller in 7-month-olds than in 4-month-olds, indicating that the older babies could either make finer perceptual discriminations or keep these in memory longer across the 1–5 item test delay. Neural systems that are present at birth are often associated with a *critical* (or *sensitive*) *period* (Sengpiel, 2007), requiring environmental input of the appropriate stimulus type within a specified period after birth to avoid being taken over for other purposes. In a classic example, cats are born with cells tuned to all line orientations, but if raised in an environment containing only vertical lines, they lose horizontal-responsive cells and demonstrate a corresponding lack of behavioral sensitivity to horizontal lines. For faces, Le Grand and colleagues report evidence consistent with a critical period for one important aspect of face perception: holistic processing. Congenital cataract patients, specifically people born with dense cataracts disrupting all pattern vision who had the cataracts removed at 2–28 months of age, were tested at ages ranging between 9 years and adulthood. Despite their many years of postcataract exposure to faces, patients who had had early bilateral cataracts showed no composite effect for faces (Le Grand et al., 2004). Also, patients who had had right-eye-only or bilateral cataracts, which produce a deficit of input to the right hemisphere due to the wiring of the infant visual system, showed a later deficit in processing spacing information in faces, while patients who had had left-eye-only cataracts did not (Le Grand, Mondloch, Maurer, & Brent, 2003), a pattern that is consistent with the normal role of the right hemisphere in holistic processing (Rossion et al., 2000). Interestingly, there does not appear to be a critical period for the ability to discriminate faces per se. Anecdotally, the Canadian cataract patients are not functionally prosopagnosic (Daphne Maurer, personal communication); for example, they report even being able to recognize other-race students when teaching English in Korea (Rachel Robbins, personal communication). Formal testing shows good ability to match novel faces (without view change) both in these patients (Geldart, Mondloch, Maurer, de Schonen, & Brent, 2002) and in an Indian woman whose congenital cataracts were not removed until 12 years of age (Ostrovsky, Andalman, & Sinha, 2006). Also, lack of visual experience with faces for the first 6–24 months in Sugita's (2008) monkeys did not destroy discrimination ability. The reason why a requirement for early visual input exists for holistic processing but not face discrimination remains to be resolved. One possibly relevant observation is that holistic processing could perhaps have a particular role in cross-view recognition (McKone, 2008), and the Canadian cataract patients have a specific problem with recognition of once-seen faces across view changes (Geldart et al., 2002). (Note that the Indian patient and Sugita's monkeys were tested on same-view faces only.)

The behavioral findings reviewed above, demonstrating abilities present at birth, perceptual narrowing and critical periods, are all consistent with a genetically determined "innate" contribution to infant face recognition. In particular, they argue for an innate contribution to face individuation.

Neurally, face individuation in adults is associated with cortical rather than subcortical function. What is the evidence regarding cortical face-processing function in infants? There are few available studies and none in neonates. Results do, however, demonstrate face selectivity and inversion effects. In infant macaques, Rodman, Scalaidhe, and Gross (1993) found that the response magnitude of single units in inferotemporal cortex was lower overall than in adults, but selectivity for form, including face selectivity, was present at the youngest ages that were tested, within 2 months of birth. In humans, a PET study of 2.5-month-olds is somewhat suggestive of face-selective activation in the fusiform gyrus (and other cortical regions), although the infants were not neurologically normal, the statistical threshold was extremely lenient ($p < 0.05$ uncorrected), and the contrast (faces versus blinking diodes) confounds selectivity for faces with responses to visual shape information (Tzourio-Mazoyer et al., 2002). With the use of ERPs, human 3-month-olds exhibit an "N290" component that has larger amplitude for human faces than for monkey faces in the right hemisphere only

(Halit, de Haan, & Johnson, 2003), although adult N170 shows the opposite pattern. At 12 months of age, this N290 was higher in amplitude for inverted faces than for upright faces only for human faces, not monkey faces (like the adult N170). Although the same study reported that this sensitivity to inversion was not found in 3-month-olds, another analysis of the same data using a different method (Johnson et al., 2005) did claim to find such inversion sensitivity. Further, other ERP components (the P400 and the P1) do show inversion effects at 3 months, the youngest age tested (Halit et al., 2003). Similarly, near-infrared spectroscopy (NIRs) responses in 5- to 8-month-old infants are stronger for upright faces than for inverted faces over the right hemisphere only (Otsuka et al., 2007; note the cortical source of this effect was most likely the STS). Overall, the available neural evidence from infants is consistent with the existence of cortical machinery for processing faces within a few months after birth, and there is no evidence to suggest that this is not present earlier.

Taking all findings together, we conclude that infants are born with a rich capacity to represent the structure of upright faces that supports face discrimination rather than merely drawing attention to faces. Results further show that this representation interacts with experience during infancy in particular ways. A probable critical period suggests that holistic processing is "experience-expectant" (i.e., early environmental input is required for its maintenance). Perceptual narrowing shows that early experience restricts the range of faces that can be accommodated; that is, an initial representation of faces is sufficiently broadly tuned to support individuation of all face types including those of other primates, and experience with one subtype of face (own-species, own-race) removes this initial ability with other face types (other-species, other-races) at the same time that it possibly improves perceptual tuning for faces of the experienced subtype. Regarding neural origin of face discrimination in infants, there is evidence of relevant cortical representation by midinfancy, but no data are available regarding whether the discrimination ability that is present at birth is supported by cortical as opposed to subcortical representations.

Development: Four-year-olds to adults

In understanding the interaction of genetic inheritance and learning, investigation of the developmental trajectory of face processing in childhood through adulthood can be informative. When no change is found in a given behavioral or neural measure of face perception in this period, that argues against extended maturation or learning as being necessary for the construction of the adult system. If instead protracted development is observed, this could reflect learning (as often assumed), though crucially it could also reflect biological maturation (Carey et al., 1980) or an interaction of genetic and experiential factors.

Behavioral Measures of Face Identity Perception For children 4–5 years and older, it is possible, with care, to adapt adult behavioral paradigms directly and thus to compare child performance with adult performance on exactly the same tasks. For each phenomenon that is established in adults, two empirical questions are of interest. First, is there some age below which children simply do not show that phenomenon at all (i.e., is there *qualitative* change with age)? Second, regarding any phenomena that are observed, when are full maturity levels reached (i.e., is there *quantitative* change with age)?

We consider qualitative change first. Early behavioral research appeared to suggest that core perceptual processes involved in face identification did not emerge at all until quite late in development (e.g., 10 years for holistic processing; Carey & Diamond, 1977; Carey et al., 1980). Unfortunately, researchers in the face neuroscience literature (e.g., Gathers, Bhatt, Corbly, Farley, & Joseph, 2004; Aylward et al., 2005; Golarai et al., 2007; Scherf, Behrmann, Humphrey, & Luna, 2007) commonly emphasize only these few early findings, which give an inaccurate representation of the current state of knowledge. In fact, research in the last 15 years has clearly established that all standard adult face recognition effects are present in young children. (Indeed, we showed earlier in the chapter that all phenomena that were tested, including inversion effects, were present in infancy.)

In child-age studies using adult tasks, every key adult property of face recognition that has been investigated has been obtained at the youngest age tested. With respect to holistic processing, these results include the inversion effect on short- and long-term recognition memory (3 years old: Sangrigoli & de Schonen, 2004; 4 years old: Carey, 1981; 5–6 years old: Brace et al., 2001; 7 years old: Flin, 1985), the composite effect (4 years old: De Heering, Houthuys, & Rossion, 2007; 6 years old: Carey & Diamond, 1994; 6 years old: Mondloch, Pathman, Maurer, Le Grand, & de Schonen, 2007), the part-whole effect for upright but not inverted faces (4 years old: Pellicano & Rhodes, 2003; 6 years old: Tanaka, Kay, Grinnell, Stansfield, & Szechter, 1998), the part-in-spacing-changed-whole effect for upright but not inverted faces (4 years old: Pellicano, Rhodes, & Peters, 2006) sensitivity to exact spacing between facial features (4 years old: McKone & Boyer, 2006; Pellicano et al., 2006), the perceptual bias to upright in superimposed faces (8 years old: Donnelly, Hadwin, Cave, & Stevenage, 2003), and the internal-over-external features advantage for familiar face identification (5–6 years old: Wilson, Blades, & Pascalis, 2007). Regarding face-space coding, results include

distinctiveness effects on perception at 4 years (McKone & Boyer, 2006) and on memory at 6–7 years (Gilchrist & McKone, 2003), an other-race disadvantage on recognition memory at 3 years (Sangrigoli & de Schonen, 2004) and a recent conference report of adaptation aftereffects in 4–5-year-olds (Jeffrey & Rhodes, 2008). Where early studies did not show effects, this has generally been established to have arisen from methodological problems, the most common one being floor effects on the task in young children (e.g., see Carey et al., 1980, versus Carey, 1981; or Johnston & Ellis, 1995, versus Gilchrist & McKone, 2003). Another case of note is the early suggestion that children could not perform face identification at all in the presence of distracting paraphernalia (Carey & Diamond, 1977); this finding was overturned (Lundy, Jackson, & Haaf, 2001) by simply making the faces larger. (Also note that even adults are sometimes strongly distracted by paraphernalia; Simons & Levin, 1998.) In summary, it is clear that there is no qualitative change in face perception beyond 4–5 years of age; quite possibly, there is none beyond infancy.

The question of whether quantitative change occurs is more difficult to answer. Certainly, performance on just about any experimental task involving faces improves very substantially across childhood and well into adolescence (see figures 32.3*A* and 32.3*B*). The crucial issue is how much of this development reflects development in face perception (e.g., in holistic processing or in the fine tuning of face-space) and how much reflects development in other general cognitive factors that are known to improve substantially across this age range and would affect task performance whatever the stimuli (e.g., explicit memory ability, ability to concentrate on the task to instruction). A common bias of face researchers is to assume, given data showing increasing memory for faces with age (e.g., figure 32.3*A*), that it is face perception that is changing, and that the task type—explicit memory—is irrelevant; yet an implicit memory researcher looking at the same set of data would likely conclude that "explicit memory" is developing and presume that the particular stimulus type—faces—is irrelevant.

Various attempts have been made to overcome the limitations of simply tracking age-related improvement in raw performance. To our minds, none of these are methodologically satisfactory, and none produce a clear conclusion regarding whether face perception per se does, or does not, improve between early childhood and adulthood. One approach is to compare two conditions across development, for example, asking whether the size of the difference between upright and inverted (or typical and distinctive, etc.) changes with age (e.g., Carey et al., 1980; Johnston & Ellis, 1995). The results of almost all such studies, however, are confounded with overall "baseline" changes across age groups, such that (1) when room to show effects is potentially compressed by approaching floor in young children but is not restricted (i.e., no ceiling effect) in adults, results seem to suggest quantitative *increases* in the effect of interest with age (figure 32.3*A*), but (2) when room to show effects is restricted by approaching ceiling in adults but is not restricted in young children (i.e., no floor effects on accuracy or, alternatively, use of a reaction time measure), results seem to show quantitative *decreases* with age (figure 32.3*B*). Taking seriously the results of the first type of study as showing quantitative development in face perception (as is commonly done) requires also taking seriously the results of the second type of study—apparently leading to the conclusion that face perception gets consistently *worse* between early childhood and adulthood! A further requirement for valid comparison of rates of development for two stimulus types is that performance be equated for the two types in one or other endpoint age group. This is commonly not done. As one example, Mondloch, Le Grand, & Maurer's (2002) finding that sensitivity to feature changes reaches adult levels earlier than spacing changes can be attributed (McKone & Boyer, 2006) simply to the fact that the features changes were easier in adults (that is, performance on an easier stimulus set reaches adult levels before performance on a more difficult stimulus set does). Another general issue in studies comparing faces versus objects, for example, in rate of development (Golarai et al., 2007) or size of inversion effects (Carey & Diamond, 1977; Teunisse & de Gelder, 2003; Aylward et al., 2005), is that in addition to producing very mixed results, the object classes that have been tested to date (houses, scenes, sculptures, shoes) have not been well matched to faces on basic parameters, such as not sharing a first-order configuration (houses, scenes) or not being natural objects (sculptures, shoes).

Overall, we conclude that current behavioral evidence demonstrates qualitatively adultlike processing of faces in young children but does not resolve whether processing is quantitatively mature. We note, however, that at least some evidence suggests a conclusion that is likely to be surprising to many readers, namely, that even quantitative maturity might be reached by early childhood. The three studies that appear to have the most suitable methodology, in which baselines were matched across age groups (Carey, 1981; Gilchrist & McKone, 2003) or restriction of range problems was otherwise avoided (Mondloch et al., 2007), all indicate no change in holistic processing (inversion effect: Carey, 1981; composite effect: Mondloch et al., 2007; spacing sensitivity: Gilchrist & McKone, 2003; or distinctiveness effects: Gilchrist & McKone, 2003) between early childhood (4–6 years) and adulthood (figure 32.3*C*).

Neural Measures of Face Identity Processing (FFA and N170)

As with behavioral studies, we discuss results of neuroimaging and ERP studies in children with respect to two issues: qualitative development and quantitative development.

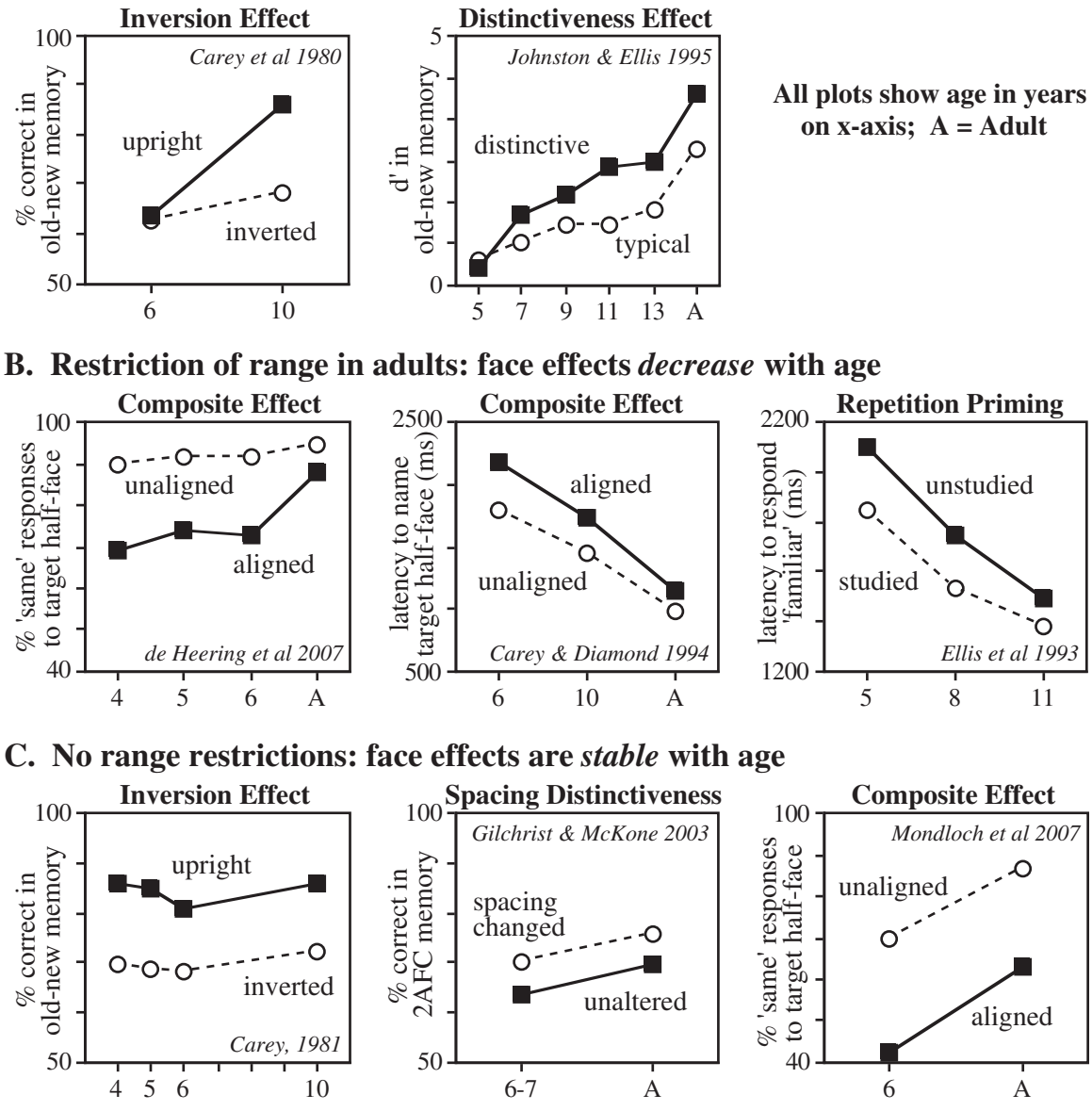


FIGURE 32.3 Behavioral face recognition effects in the preschooler to adult age range. A basic finding is of overall improvement with age: higher accuracy or lower reaction time. Note that in part *C*, the left and middle plots show studies in which the researchers deliberately removed this trend by using smaller learning set sizes in younger children. Our major point is that apparent developmental trends in the strength of core effects (size of inversion effect, size of composite effect, ability to represent recently seen faces in implicit memory, etc.) depend on whether and how room to show effects is potentially restricted.

Three studies have used fMRI to scan children age 5 years to adult on face and object tasks, enabling these studies to track the existence and size of face-selective regions of cortex (figure 32.4). (A fourth study will not be discussed here because it used such liberal criteria to define "FFAs" that the regions that were so identified were clearly not face-selective even in adults; see figure 1d–f in that study, Gathers et al., 2004.) Considering qualitative effects, evidence of a face-selective FFA has been found in most children at the youngest ages tested. Although no FFA was revealed in young children by group analyses (in which all subjects are aligned in a common space; 5- to 8-years old: Scherf et al., 2007; 8–10 years old: Aylward et al., 2005), in the two studies reporting individual-subject analyses, Scherf and colleagues found an FFA in 80% of the children in 5- to 8-year-olds (albeit at a very liberal statistical threshold), and Golarai and colleagues (2007) found an FFA in 85% of children in their 7- to 11-year-old group (using a more standard statistical threshold). One study (Passarotti, Smith, DeLano, & Huang, 2007) also reported an inversion effect (a higher response to inverted

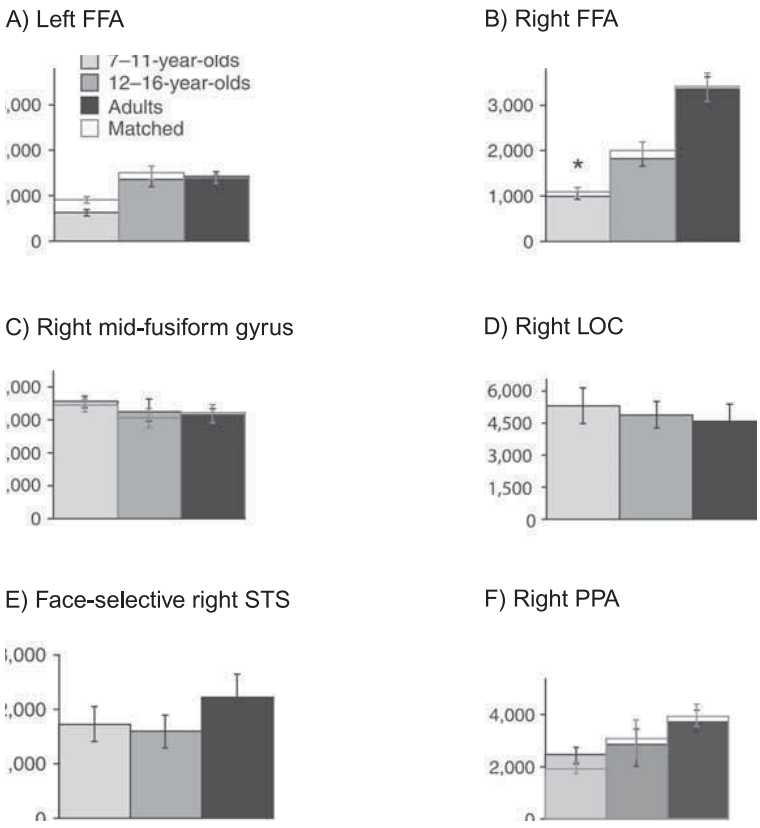


FIGURE 32.4 Mean volume across subjects in each age group of individually defined (*A*) left and (*B*) right FFA, (*C*) anatomically defined right mid-fusiform gyrus, (*D*) functionally defined right LOC, and functionally defined (*E*) face-selective right STS and (*F*) right place-selective PPA. Red bars indicated values in subsets of subjects matched for BOLD-related confounds. (From Golarai et al., 2007.) (See color plate 47.)

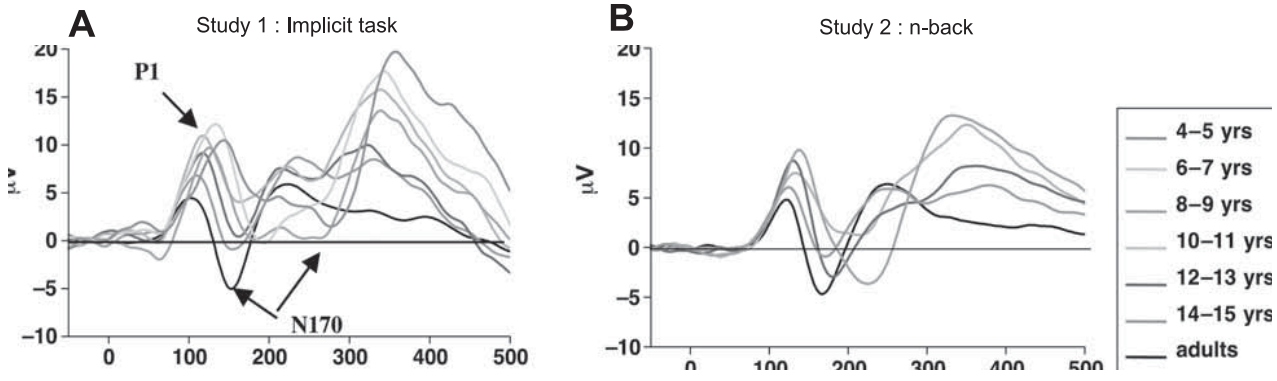


FIGURE 32.5 ERPs from right posterior temporal scalp locations in response to face stimuli, separately for each age group. (From Taylor et al., 2004.) (See color plate 48.)

faces than to upright faces) in the region of the right (but not the left) FFA in children 8–11 years of age (and an effect in the opposite direction in adults). Regarding ERPs, young children (like infants) show both face-selective responses and inversion effects upon these (see figures 32.5 and 32.6; Taylor, Batty, & Itier, 2004). These fMRI and ERP findings in children add to the infant data to confirm that at least some form of face-specific neural machinery is established early.

Quantitatively, the neural machinery that is involved in face perception demonstrates substantial changes in face-selective neural responses continuing late into development. In all three fMRI studies, the FFA increases markedly in volume between childhood and adulthood (Aylward et al., 2005; Golarai et al., 2007; Scherf et al., 2007), even though total brain volume does not change substantially after age 5 years. These studies clearly show that the rFFA is still chang-

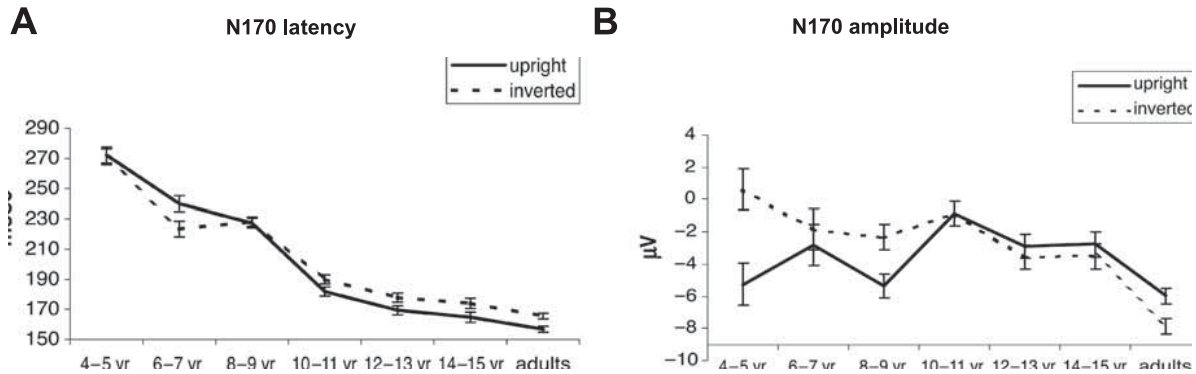


FIGURE 32.6 Mean N170 latency (left) and amplitude (right) for upright and inverted faces as a function of age. (From Taylor et al., 2004.)

ing late in life—certainly after age 7 and in some studies much later.

Comparing fMRI data across children and adults is fraught with potential pitfalls. Children move more in the scanner and are less able to maintain attention on a task. These or other differences between children and adults could in principle explain the change in volume of the rFFA. However, notably, control areas that are identified in the same scanning sessions do not change with age. For example, object-responsive regions and the scene-selective "parahippocampal place area" in the right hemisphere or rPPA (Epstein & Kanwisher, 1998) did not change in volume from childhood to adulthood (Golarai et al., 2007; Scherf et al., 2007), although somewhat surprisingly, Golarai and colleagues found that the lPPA did increase in volume with age. These findings reassure us that the changes in the rFFA with age are not due to across-the-board changes in the ability to extract good functional data from young children.

Golarai and colleagues (2007) asked how changes in the rFFA relate to changes in behavioral face recognition over development. Right FFA size was correlated (separately in children and adolescents but not in adults) with face recognition memory but not with place or object memory. Conversely, lPPA size was correlated (in all age groups independently) with place memory but not with object or face memory. This double dissociation of behavioral correlations clearly associates the rFFA with changes in face recognition measured behaviorally.

ERP findings are consistent with the evidence from fMRI that the cortical regions that are involved in face recognition continue to change well into the teenage years. Face-related ERPs show gradual changes in scalp distribution, latency, and amplitude into the mid-teen years (figures 32.5 and 32.6). Both the early P1 component and the later N170 component show gradual decreases in latency from age 4 to adulthood. Regarding neural inversion effects, late developmental changes are found with both fMRI and ERP (see figure 32.5), including a reversal of the direction of the inversion effect between children and adults in both methods (Taylor et al., 2004; Passarotti et al., 2007). Future research might best approach this question not just by measuring mean responses to upright versus inverted faces, but also by using identity-specific adaptation to ask when the better discrimination of upright than inverted faces seen in adulthood emerges (Yovel & Kanwisher, 2005; Mazard et al., 2006).

COMPARING DEVELOPMENT FOR BEHAVIORAL AND NEURAL MEASURES Taking the findings from the 4-to-adult range together with the infant literature, we can draw the following conclusions. First, the results regarding qualitatively adultlike face processing appear to agree well across behavioral and neural measures; that is, just as all behavioral face recognition effects have been obtained in the youngest age groups tested, face-selective neural machinery as revealed by fMRI, ERPs, NIRs, and single-cell recording has also been found in the youngest children and infants tested. Nonetheless, fMRI data are not available for children younger than 5–8 (pooled together), and the ERP studies in infants and children often go in opposite directions from those in adults. For example, the inversion effect on the N170 switches polarity between childhood and adulthood, as shown in figure 32.6, despite maintaining the same polarity in behavior.

Second, the evidence for quantitative development is less clear. It might be that the improvements with age on behavioral tasks do reflect ongoing development of face perception itself; if so, this could agree neatly with the increasing size of the FFA. As we have noted, however, findings such as those shown in figures 32.3*B* and 32.3*C* suggest that behavioral face perception could be fully mature early and that ongoing behavioral improvements with age reflect changes in other, more general, cognitive factors. This view would produce an apparent discrepancy—behavioral maturity arising well before maturity of relevant cortical regions—that would need to be resolved. If this is the case, two ideas

	Newborns / deprived	< = 3 months	Later infancy	4 years	5 years	6 years	7 years	8 years	9 years	10 years	11 years	Adults
Behavioral Properties												
Ability to discriminate individual faces	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺
Inversion effect on discrimination (looking time or recognition memory)	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺
Composite-like effect, upright not inverted			☺									☺
Composite effect				☺	☺	☺	☺	☺	☺	☺	☺	☺
Part-whole effect, upright not inverted				☺	☺	☺	☺	☺	☺	☺	☺	☺
Part-in-spacing-altered-whole effect, upright not inverted				☺	☺	☺	☺	☺	☺	☺	☺	☺
Sensitivity to spacing changes	☺		☺	☺	☺	☺	☺	☺	☺	☺	☺	☺
Inversion effect on spacing sensitivity			☺			☺	☺	☺	☺	☺	☺	☺
Perceptual bias to upright in superimposed faces							☺	☺	☺			☺
Distinctiveness effects				☺	☺	☺	☺	☺	☺	☺	☺	☺
Adaptation aftereffects				☺	☺				☺	☺	☺	☺
Attractiveness preference, upright not inverted	☺											☺
Neural Properties												
Face-selective cells, macaques		☺										☺
Face-selective ERPs		☺		☺	☺	☺	☺	☺	☺	☺	☺	☺
FFA present					☺	☺	☺	☺	☺	☺	☺	☺
Some type of inversion effect on neural response		?	☺	☺	☺	☺	☺	☺	☺	☺	☺	☺
Perceptual Narrowing												
Looking time discrimination of other race/species faces	☺	☺	X									X

FIGURE 32.7 For each property of face processing, we indicate for each age group whether that property is *qualitatively* present (☺), debatable (?), not present (X), or not yet tested (gray). Deprived = monkeys deprived of face input from birth. Note: All references can be found in text except inversion effect on spacing sensitivity aged 6 years to adult is from Mondloch et al. (2002) and adaptation aftereffect aged 9 years to adult is from Pellicano, Jeffery, Burr, & Rhodes (2007).

might be worth exploring. It might be that the measured size of the FFA in children is affected by top-down strategic processing that (for some unknown reason) affects faces and not objects. Another possibility is that the FFA might play some role in the long-term storage of individual faces (e.g., it shows repetition priming; Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005; Williams, Berberovic, & Mattingley, 2007) and that the increased size of the FFA could arise simply because people continue to learn faces across life; this idea would have to propose that the number of new faces learned is much greater than the number of new objects.

Conclusion

For decades, conventional wisdom has held that face recognition arises very slowly in development and that experience is the primary engine of this development. The new evidence that we have reviewed here refutes this hypothesis. Impressive face recognition abilities are present within a few days of birth and are present in monkeys who have never seen faces before. Some form of inherited genetic influence is also indicated by Polk and colleagues' imaging study of twins and by the fact that developmental prosopagnosia can run in families. Qualitatively, behavioral findings indicate establishment of all adultlike face recognition effects by 4 years at the latest and in infancy wherever tested; the striking breadth of this evidence is summarized in figure 32.7. The available evidence also indicates early initial establishment of face-selective neural machinery at the cortical level (again see figure 32.7). It is not, however, that experience plays no role in development. Perceptual narrowing of the range of facial subtypes for which discrimination is possible reveals a destructive role for experience. Further, there is a requirement for early-infancy input (consistent with a critical period) for the development of holistic face processing but (mysteriously) not face discrimination.

Three major questions remain for future research. First, it will be critical to determine whether face perception per se improves quantitatively after age 4 years or whether instead improvement in performance after this age reflects improvement in domain-general mechanisms. Second, if face perception itself does improve quantitatively after age 4, what role does experience play in this improvement? A final critical challenge will be to understand the relationship between cognitive and neural development, especially the substantial increase in the size of the FFA.

ACKNOWLEDGMENTS We thank Mark Johnson for useful comments and Bettiann McKay for help with the manuscript. Preparation of this chapter was supported by NEI grant 13455 and a grant from the Ellison Foundation to N.K. and by grant DP0450636 from the Australian Research Council to E.M.

REFERNCES

Aylward, E. H., Park, J. E., Field, K. M., Parsons, A. C., Richards, T. L., Cramer, S. C., & Meltzoff, A. N. (2005). Brain activation during face perception: Evidence of a developmental change. *J. Cogn. Neurosci.*, *17*(2), 308–319.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *J. Cogn. Neurosci.*, *8*(6), 551–565.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychol. Sci.*, *8*, 706–712.

Brace, N. A., Hole, G. J., Kemp, R. I., Pike, G. E., Van Duuren, M., & Norgate, L. (2001). Developmental changes in the effect of inversion: Using a picture book to investigate face recognition. *Perception*, *30*, 85–94.

Bruce, V., Henderson, Z., Newman, C., & Burton, M. A. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *J. Exp. Psychol. [Appl.]*, *7*(3), 207–218.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychol. Sci.*, *10*(3), 243–248.

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant Child Dev.*, *10*, 67–74.

Carbon, C. C., Strobach, T., Langton, S., Harsányi, G., Leder, H., & Kovács, G. (2007). Adaptation effects of highly familiar faces: Immediate and long lasting. *Mem. Cogn.*, *35*(8), 1966–1976.

Carey, S. (1981). The development of face perception. In G. Davies, H. D. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 9–38). New York: Academic Press.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, *195*, 312–314.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cogn.*, *2/3*, 253–274.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Dev. Psychol.*, *16*(4), 257–269.

Cohen, L. B., & Cashon, C. H. (2001). Do 7-month-old infants process independent features of facial configurations? *Infant Child Dev.*, *10*, 83–92.

de Haan, M., Johnson, M. H., Maurer, D., & Perrett, D. I. (2001). Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. *Cogn. Dev.*, *16*, 659–678.

de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *J. Exp. Child Psychol.*, *96*, 57–70.

DeGutis, J. M., Bentin, S., Robertson, L. C., & D'Esposito, M. (2007). Functional plasticity in ventral temporal cortex following cognitive rehabilitation of a congenital prosopagnosic. *J. Cogn. Neurosci.*, *19*(11), 1790–1802.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.*, *4*(8), 2051–2062.

Donnelly, N., Hadwin, J. A., Cave, K., & Stevenage, S. (2003). Perceptual dominance of oriented faces mirrors the distribution of orientation tuning in inferotemporal neurons. *Cogn. Brain Res.*, *17*, 771–780.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cogn. Neuropsychol.*, *24*(4), 419–430.

Duchaine, B. C., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*, *41*(6), 713–720.

Ellis, H. D., Sheperd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*(4), 431–439.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proc. Natl. Acad. Sci. USA*, *102*(47), 17245–17250.

Flin, R. H. (1985). Development of face recognition: An encoding switch? *Br. J. Psychol.*, *76*, 123–134.

Foldiak, P., Xiao, D., Keysers, C., Edwards, R., & Perrett, D. I. (2004). Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Prog. Brain Res.*, *144*, 107–116.

Gathers, A. D., Bhatt, R. S., Corbly, C. R., Farley, A. B., & Joseph, J. E. (2004). Developmental shifts in cortical loci for face and object recognition. *NeuroReport*, *15*(10), 1549–1553.

Geldart, S., Mondloch, C. J., Maurer, D., de Schonen, S., & Brent, H. P. (2002). The effect of early visual deprivation on the development of face processing. *Dev. Sci.*, *5*(4), 490–501.

Gilchrist, A., & McKone, E. (2003). Early maturity of face processing in children: Local and relational distinctiveness effects in 7-year-olds. *Visual Cogn.*, *10*(7), 769–793.

Golarai, G., Ghahemani, D. G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J. L., Gabrieli, J. D., & Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat. Neurosci.*, *10*, 512–522.

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*(4), 544–549.

Grueter, M., Grueter, T., Bell, V., Horst, J., Laskowski, W., Sperling, K., Halligan, P. W., Ellis, H. D., & Kennerknecht, I. (2007). Hereditary prosopagnosia: The first case series. *Cortex*, *43*(6), 734–749.

Halgren, E., Raij, T., Marinkovic, K., Jousmaki, V., & Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb. Cortex*, *10*(1), 69–81.

Halit, H., de Haan, M., & Johnson, M. H. (2003). Cortical specialisation for face processing: Face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*, *19*(3), 1180–1193.

Hayden, A., Bhatt, R. S., Reed, A., Corbly, C. R., & Joseph, J. E. (2007). The development of expert face processing: Are infants sensitive to normal differences in second-order relational information? *J. Exp. Child Psychol.*, *97*, 85–98.

Humphreys, K., & Johnson, M. H. (2007). The development of "face-space" in infancy. *Visual Cogn.*, *15*(5), 578–598.

Jacques, C., d'Arripe, O., & Rossion, B. (2007). The time course of the inversion effect during individual face discrimination. *J. Vis.*, *7*(8), 1–9.

Jacques, C., & Rossion, B. (2006). The speed of individual face categorization. *Psychol. Sci.*, *17*(6), 485–492.

Jeffrey, L., & Rhodes, G. (2008). *Aftereffects reveal enhanced face-coding plasticity in young children.* Poster presented at meeting of Vision Sciences Society, Naples, FL.

Johnson, M. H. (2005). Subcortical face processing. *Nat. Rev. Neurosci.*, *6*, 766–774.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1–2), 1–19.

Johnson, M. H., Griffin, R., Csibra, G., Halit, H., Farroni, T., de Haan, M., Tucker, L. A., Baron-Cohen, S., & Richards, J. (2005). The emergence of the social brain network: Evidence from typical and atypical development. *Dev. Psychopathol.*, *17*(3), 599–619.

Johnston, R. A., & Ellis, H. D. (1995). Age effects in the processing of typical and distinctive faces. *Q. J. Exp. Psychol. [A]*, *48*(2), 447–465.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.*, *17*(11), 4302–4311.

Kanwisher, N., & Yovel, G. (in press). Cortical specialization for face perception in humans. In J. T. Cacioppo & G. G. Berntson (Eds.), *Handbook of neuroscience for the behavioral sciences*. Hoboken, NJ: John Wiley.

Kanwisher, N. G., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.*, *17*(11), 4302–4311.

Kelly, D. J., Quinn, P. C., Slater, A., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychol. Sci.*, *18*(12), 1084–1089.

Kennerknecht, I., Pluempe, N., & Welling, B. (2008). Congenital prosopagnosia: A common hereditary cognitive dysfunction in humans. *Front. Biosci.*, *1*(13), 3150–3158.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritanim, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.*, *9*(2), F13–F21.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. USA*, *100*, 9096–9101.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nat. Neurosci.*, *6*(10), 1108–1112.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychol. Sci.*, *15*(11), 762–768.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*(7102), 572–575.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.*, *4*(1), 89–94.

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proc. Natl. Acad. Sci. USA*, *103*(17), 6771–6774.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nat. Neurosci.*, *5*(9), 910–916.

Logie, R. H., Baddeley, A. D., & Woodhead, M. M. (1987). Face recognition, pose and ecological validity. *Appl. Cogn. Psychol.*, *1*, 53–69.

Lundy, B. L., Jackson, J. W., & Haaf, R. A. (2001). Stimulus properties, attentional limitations, and young children's face recognition. *Percept. Mot. Skills*, *92*, 919–929.

Martini, P., McKone, E., & Nakayama, K. (2006). Orientation tuning of human face processing estimated by contrast matching on transparency displays. *Vis. Res.*, *46*(13), 2102–2109.

Maurer, D., Lewis, T. L., & Mondloch, C. J. (2005). Missing sights: Consequences for visual cognitive development. *Trends Cogn. Sci.*, *9*(3), 144–151.

Mazard, A., Schiltz, C., & Rossion, B. (2006). Recovery from adaptation to facial identity is larger for upright than inverted faces in the human occipito-temporal cortex. *Neuropsychologia*, *44*(6), 912–922.

McCarthy, G., Luby, M., Gore, J., & Goldman-Rakic, P. (1997). Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. *J. Neurophysiol.*, *77*(3), 1630–1634.

McKone, E. (2008). Configural processing and face viewpoint. *J. Exp. Psychol. Hum. Percept. Perform.*, *34*(2), 310–327.

McKone, E., Aitkin, A., & Edwards, M. (2005). Categorical and coordinate relations in faces, or Fechner's law and face space instead? *J. Exp. Psychol. Hum. Percept. Perform.*, *31*(6), 1181–1198.

McKone, E., & Boyer, B. L. (2006). Sensitivity of 4-year-olds to featural and second-order relational changes in face distinctiveness. *J. Exp. Child Psychol.*, *94*(2), 134–162.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends Cogn. Sci.*, *11*(1), 8–15.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychol. Public Policy Law*, *7*(1), 3–35.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, *31*, 553–566.

Mondloch, C. J., Pathman, T., Maurer, D., Le Grand, R., & de Schonen, S. (2007). The composite face effect in six-year-old children: Evidence of adult-like holistic face processing. *Visual Cogn.*, *15*(5), 564–577.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychol. Rev.*, *98*(2), 164–181.

Ostrovsky, Y., Andalman, A., & Sinha, P. (2006). Vision following extended congenital blindness. *Psychol. Sci.*, *17*(12), 1009–1014.

Otsuka, Y., Nakato, E., Kanazawa, S., Yamaguchi, M. K., Watanabe, S., & Kakigi, R. (2007). Neural activation to upright and inverted faces in infants measured by near infrared spectroscopy. *NeuroImage*, *34*, 399–406.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*, 1321–1323.

Pascalis, O., de Haan, M., Nelson, C. A., & de Schonen, S. (1998). Long-term recognition memory for faces assessed by visual paired comparison in 3- and 6-month-old infants. *J. Exp. Psychol. Learn. Mem. Cogn.*, *24*(1), 249–260.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behav. Dev.*, *18*, 79–85.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proc. Natl. Acad. Sci. USA*, *102*(14), 5297–5300.

Passarotti, A. M., Smith, J., DeLano, M., & Huang, J. (2007). Developmental differences in the neural bases of the face inversion effect show progressive tuning of face-selective regions to the upright orientation. *NeuroImage*, *34*(4), 1708–1722.

Pellicano, E., Jeffrey, L., Burr, D., & Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Curr. Biol.*, *17*(17), 1508–1542.

Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychol. Sci.*, *14*(6), 618–622.

Pellicano, E., Rhodes, G., & Peters, M. (2006). Are preschoolers sensitive to configural information in faces? *Dev. Sci.*, *9*(3), 270–277.

Perrett, D. I., Mistlin, A. J., Chitty, A. J., Smith, P. A., Potter, D. D., Broennimann, R., & Harries, M. (1988). Specialized face processing and hemispheric asymmetry in man and monkey: Evidence from single unit and reaction time studies. *Behav. Brain Res.*, *29*(3), 245–258.

Pinsk, M. A., DeSimone, K., Moore, T., Gross, C.G., & Kastner, S. (2005). Representations of faces and body parts in macaque temporal cortex: A functional MRI study. *Proc. Natl. Acad. Sci., USA*, *102*(19), 6996–7001.

Polk, T. A., Park, J., Smith, M. R., & Park, D. C. (2007). Nature versus nurture in ventral visual cortex: A functional magnetic resonance imaging study of twins. *J. Neurosci.*, *27*(51), 13921–13925.

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2005). Portraits or people? Distinct representations of face identity in the human visual cortex. *J. Cogn. Neurosci.*, *17*(7), 1043–1057.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *J. Neurosci.*, *16*(16), 5205–5215.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition*, *47*, 25–57.

Robbins, R., & McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition*, *88*, 79–107.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*(1), 34–79.

Rodman, H. R., Scalaidhe, S. P., & Gross, C. G. (1993). Response properties of neurons in temporal cortical visual areas of infant monkeys. *J. Neurophysiol.*, *70*(3), 1115–1136.

Rossion, B., Dricot, L., A., D., Bodart, J., Crommelinck, M., de Gelder, B., & Zoontjes, R. (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *J. Cogn. Neurosci.*, *14*(5), 793–802.

Rotshtein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat. Neurosci.*, *8*(1), 107–113.

Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant Child Dev.*, *14*, 29–50.

Sangrigoli, S., & de Schonen, S. (2004). Effect of visual experience on face processing: A developmental study of inversion and non-native effects. *Dev. Sci.*, *7*(1), 74–87.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychol. Sci.*, *16*(6), 440–444.

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30*(4), 1088–1096.

Scherf, K. S., Behrmann, M., Humphrey, K., & Luna, B. (2007). Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Dev. Sci.*, *10*(4), F15–F30.

Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *NeuroImage*, *32*, 1385–1394.

Sengpiel, F. (2007). The critical period. *Curr. Biol.*, *17*, R742–R743.

Simion, F., Macchi Cassia, V., Turati, C., & Valenza, E. (2003). Non-specific perceptual biases at the origins of face processing. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood: Current perspectives* (pp. 13–25). New York: Nova Science.

Simion, F., Valenza, E., Umilta, C., & Dalla Barba, B. (1998). Preferential orienting to faces in newborns: A temporal-nasal

asymmetry. *J. Exp. Psychol. Hum. Percept. Perform.*, *24*(5), 1399–1405.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during real-world interaction. *Psychon. Bull. Rev.*, *5*(4), 644–649.

Slater, A., Quinn, P. C., Hayes, R., & Brown, E. (2000). The role of facial orientation in newborn infants' preference for attractive faces. *Dev. Sci.*, *3*(2), 181–185.

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proc. Natl. Acad. Sci. USA*, *105*(1), 394–398.

Tanaka, J. W., & Farah, M. (1993). Parts and wholes in face recognition. *Q. J. Exp. Psychol.*, *46A*(2), 225–245.

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cogn.*, *5*, 479–496.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Mem. Cogn.*, *25*(5), 583–592.

Taylor, M. J., Batty, M., & Itier, R. J. (2004). The faces of development: A review of early face processing over childhood. *J. Cogn. Neurosci.*, *16*(8), 1426–1442.

Teunisse, J. P., & de Gelder, B. (2003). Face processing in adolescents with autistic disorder: The inversion and composite effects. *Brain Cogn.*, *52*(3), 285–294.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nat. Neurosci.*, *6*(9), 989–995.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*, 670–674.

Turati, C., Bulf, H., & Simion, F. (2008). Newborns' face recognition over changes in viewpoint. *Cognition*, *106*, 1300–1321.

Turati, C., Macchi Cassia, V., Simion, F., & Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. *Child Dev.*, *77*(2), 297–311.

Tzourio-Mazoyer, N., de Schonen, S., Crivello, F., Reutter, B., Aujard, Y., & Mazoyer, B. (2002). Neural correlates of woman face processing by 2-month-old infants. *NeuroImage*, *15*, 454–461.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol.*, *43A*(2), 161–204.

Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception*, *15*, 525–535.

Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychon. Bull. Rev.*, *6*(4), 647–653.

Williams, M. A., Berberovic, N., & Mattingley, J. B. (2007). Abnormal fMRI adaptation to unfamiliar faces in a case of developmental prosopamnesia. *Curr. Biol.*, *17*(14), 1259–1264.

Wilson, R. R., Blades, M., & Pascalis, O. (2007). What do children look at in an adult face with which they are personally familiar? *B. J. Dev. Psychol.*, *25*, 375–382.

Winston, J. S., Vuilleumier, P., & Dolan, R. J. (2003). Effects of low-spatial frequency components of fearful faces on fusiform cortex activity. *Curr. Biol.*, *13*(20), 1824–1829.

Yin, R. K. (1969). Looking at upside-down faces. *J. Exp. Psychol.*, *81*(1), 141–145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747–759.

Yovel, G., & Duchaine, B. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. *J. Cogn. Neurosci.*, *18*(4), 580–593.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Curr. Biol.*, *15*(24), 2256–2262.

33 Roles of Visual Area MT in Depth Perception

GREGORY C. DEANGELIS

ABSTRACT One of the most impressive capacities of the visual system is the ability to infer the three-dimensional structure of the environment from images formed on the two retinas. Although several areas of visual cortex are involved in computing depth, the precise roles of different areas in three-dimensional vision remain unclear. It is important to establish how neural representations of depth in different brain regions are specialized to perform different tasks. This chapter summarizes studies that establish such links between representation and function in visual area MT. The nature of the representation of binocular disparity in MT is first considered, along with the functional roles of MT in coarse and fine depth discrimination. The recently discovered role of area MT in computing depth from motion parallax is then examined. These findings are compared with those from other visual areas to consider possible functional streams of analysis in three-dimensional vision.

We carry out our daily activities in a three-dimensional (3D) environment. Therefore a fundamental task for the visual system is to construct a 3D representation of our surroundings. This is difficult because the image formed on the retina of each eye is a two-dimensional projection of 3D space—hence there is no direct quantitative information about depth in a single retinal image. Rather, the depth structure of the scene must be reconstructed by the brain.

The visual system makes use of a wide variety of cues to estimate depth relationships (Howard & Rogers, 1995, 2002). Broadly speaking, these cues can be placed into two categories that I shall label *pictorial cues* and *geometric cues*. Pictorial cues to depth are those that are present in a single snapshot of the scene, including occlusion, perspective, shading, relative size, texture gradients, and blur. Together, these cues can be potent, as is evidenced by the fact that we can infer depth relationships in photographs. However, they generally provide only ordinal depth information or require prior knowledge to provide metric depth information. For example, the size of an object in the retinal image can be used to estimate the distance to that object if one knows the true physical size of the object.

GREGORY C. DEANGELIS Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York

Geometric depth cues are those that arise when a scene is viewed from multiple vantage points. For species with frontally located eyes, the horizontal separation between the two eyes generates systematic differences—known as binocular disparities—between the images projected onto the two retinas (figure 33.1*A*). Thus images from two simultaneous vantage points are available. Binocular disparity (hereafter referred to as *disparity*) is known to be sufficient to provide precise depth discrimination in the absence of other depth cues, as demonstrated with random-dot stereograms (Howard & Rogers, 1995; Julesz, 1971; Parker, 2007). Combined with an estimate of viewing distance (the distance from the eye to the point of fixation), disparity can provide quantitative estimates of the location of objects in depth. Another geometric cue, motion parallax, arises because of the translation of the observer, as illustrated in figure 33.1*B*. As the observer's head moves from left to right, for example, the vantage point of the left eye changes over time. If the observer's head moves through one interocular distance, then the image that is projected onto the retina of the left eye will vary over time, the endpoint being the same view that would be seen by the right eye at the beginning of the movement (figure 33.1*B*). Thus there is a formal geometric similarity between disparity and motion parallax cues, at least when the latter arise because of lateral head movements. This means that motion parallax can provide metric depth information when a subject views the scene with one eye, as long as the eye moves relative to the scene. Not surprisingly, then, humans can make judgments of depth from motion parallax that are similar in precision to judgments based on disparity (Rogers & Graham, 1979, 1982). As we shall see later, the similar geometry of these two cues suggests that they might be processed using the same neural mechanisms.

Where and how are depth cues processed in the brain? For the pictorial depth cues, very little is known about the neural mechanisms that lead to depth percepts; therefore I shall not consider pictorial cues further here. Until recently, very little was also known about the neural basis of depth from motion parallax, and we shall consider the available physiological information in the last section of this chapter. By comparison, a great deal is known about the neural circuits that process disparity cues for depth perception, as has

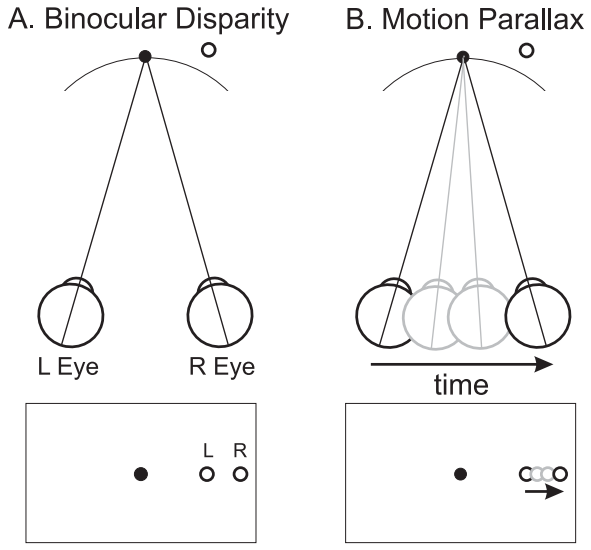


FIGURE 33.1 Binocular disparity and motion parallax as depth cues. (*A*) Points falling along the geometric horopter, or Vieth-Muller circle (curved line), have zero binocular disparity. A far object (open symbol) projects to disparate points in the two retinal images (bottom). (*B*) If the head translates rightward, the image of the far object moves on the retina. If the eye moves through one interocular distance, the position change on the retina due to motion parallax is equivalent to the binocular disparity. Hence depth from motion parallax is often expressed in units of equivalent disparity.

also been reviewed elsewhere (Cumming & DeAngelis, 2001; DeAngelis, 2000; Gonzalez & Perez, 1998; Parker, 2007). I shall focus mainly on work that has been performed using macaque monkeys as experimental subjects. However, it should be noted that many important contributions have also been made in other species, particularly cats (for reviews, see Freeman, 2004; Freeman & Ohzawa, 1990; Ohzawa, DeAngelis, & Freeman, 1997).

The primary visual cortex (V1) was the major focal point of most early physiological studies of disparity processing (Barlow, Blakemore, & Pettigrews, 1967; Poggio & Fischer, 1977). Neurons in V1 provide the initial encoding of disparity signals (Cumming & DeAngelis, 2001), and the representation of disparities in V1 likely limits the precision of a number of aspects of stereopsis (Nienborg, Bridge, Parker, & Cumming, 2004, 2005; Prince, Pointon, Cumming, & Parker, 2000). However, a series of elegant experiments over the past decade has demonstrated that the representation of disparity in V1 is not sufficient to account for various aspects of our perceptual experience of depth (Cumming & Parker, 1999, 2000, 1997; Nienborg & Cumming, 2006). Therefore it seems clear that disparity processing beyond V1 is critical to account for behavior. During this same period, we have learned that disparity signals are represented in a variety of visual cortical areas that were not previously known to contain disparity-selective neurons, including area V4 (Hegde & Van Essen, 2005; Hinkle & Connor, 2005; Tanabe, Doi, Umeda, & Fujita, 2005), inferotemporal (IT) cortex (Janssen, Vogels, Liu, & Orban, 2003; Janssen, Vogels, & Orban, 1999, 2000b; Uka, Tanaka, Yoshiyama, Kato, & Fujita, 2000), and lateral area MST (Eifuku & Wurtz, 1999). Disparity selective-neurons have also been documented in areas not conventionally considered to be predominantly visual, such as the frontal eye fields (Ferraina, Pare, & Wurtz, 2000) and the lateral (Genovesio & Ferraina, 2004; Gnadt & Mays, 1995) and caudal (Taira, Tsutsui, Jiang, Yara, & Sakata, 2000; Tsutsui, Jiang, Yara, Sakata, & Taira, 2001; Tsutsui, Sakata, Naganuma, & Tanaka, 2002) portions of the intraparietal sulcus. Combined with earlier studies showing the presence of disparity-tuned neurons in areas V1, V2, V3, V3A, MT, and dorsal MST (reviewed in Cumming & DeAngelis, 2001), these studies suggest that disparity signals are much more widely distributed in the primate brain than was suspected a decade ago.

The proliferation of disparity signals in visual cortex raises fundamental questions regarding the roles of different cortical areas: How are the representations of disparity specialized in these different areas, and what are the specific roles that particular areas play in 3D vision? These are not simple questions to answer because they require both a detailed understanding of the responses of single neurons to a variety of stimulus configurations and causal tests in which neural activity is manipulated and the consequences on behavior are observed. Over the past several years, considerable progress has been made on these fronts, and there is cause to be optimistic that these experimental approaches will lead to a functional taxonomy that describes how different visual areas contribute to the various perceptual capacities of depth perception.

My goal in this chapter is to summarize what we know about how disparity and motion parallax cues are processed in the middle temporal (MT) area of visual cortex and to review experiments that suggest a specialized functional role for area MT in depth perception. Along the way, I shall try to place the findings from area MT into the context of what is known from related studies performed in other parts of primate visual cortex.

Binocular disparity processing in area MT

Area MT is a relatively small (~60 mm^2) visual area located along the posterior bank of the superior temporal sulcus in macaque monkeys. It receives much of its visual input from areas V1, V2, and V3, and it projects extensively to other occipitoparietal areas, including MST, FST, VIP, and LIP (reviewed in Born & Bradley, 2005). Area MT is well known for its role in processing visual motion, and an extensive body of literature implicates area MT both in the perception of motion and in guiding smooth eye movements that are driven by visual image motion (Born & Bradley, 2005).

Basic Aspects of Disparity Processing in MT Despite heavy emphasis on the role of MT in motion processing, it has been known for quite some time that MT is rich in neurons that are selective for binocular disparities (Maunsell & Van Essen, 1983b). A more recent quantitative study, using random-dot stereograms (figure 33.2*A*), reported that 93% of single neurons in MT have statistically significant selectivity for disparity (DeAngelis & Uka, 2003). The strength of disparity selectivity in MT is, on average, stronger than that seen in areas V1 (Prince, Pointon, Cumming, & Parker, 2002) and V4 (Tanabe et al., 2005). As shown in figure 33.2*B*, disparity tuning curves in MT take on a variety of shapes and have continuously varying preferences over a wide range of disparities (note that zero disparity represents a surface containing the point of ocular fixation). Most frequently, however, MT neurons tend to have tuning that is roughly odd-symmetric around zero disparity, with a well-defined peak response that is either near (e.g., cells 2 and 3 in figure 33.2*B*) or far (cells 5 and 6). In this regard, MT differs notably from area V1, where disparity tuning curves are much more frequently even-symmetric and tend to peak at disparities closer to zero (Cumming & DeAngelis, 2001). As one ascends the dorsal visual pathway in the macaque, disparity tuning curves progress from largely even-symmetric in V1 to having an odd-symmetric bias in MT to being strongly odd-symmetric in area MST (see Cumming & DeAngelis, 2001, for comparison; Roy, Komatsu, & Wurtz, 1992; Takemura, Inoue, Kawano, Quaia, & Miles, 2001).

Nearly all neurons in area MT show directionally selective responses to visual motion (DeAngelis & Uka, 2003; Maunsell & Van Essen, 1983a; Zeki, 1974), and nearly all neurons are selective for speed of motion as well (Maunsell & Van Essen, 1983a; Nover, Anderson, & DeAngelis, 2005; Rodman & Albright, 1987). Thus most MT neurons respond substantially more strongly to a moving stimulus than a stationary one. Importantly, however, the majority of MT neurons do produce sustained visual responses to stationary flashed stimuli, and the disparity tuning of these responses is nearly identical to those elicited by moving stimuli (Palanca & DeAngelis, 2003). Thus MT provides reliable disparity signals to the rest of the brain when we look around a stationary scene.

Disparity-selective neurons in area MT are organized into a topographic map according to their disparity preferences (DeAngelis & Newsome, 1999), and this disparity map coexists with the well-known map for direction of motion in MT (Albright, Desimone, & Gross, 1984). In contrast to the clear columnar organization for disparity in MT, there is no clear

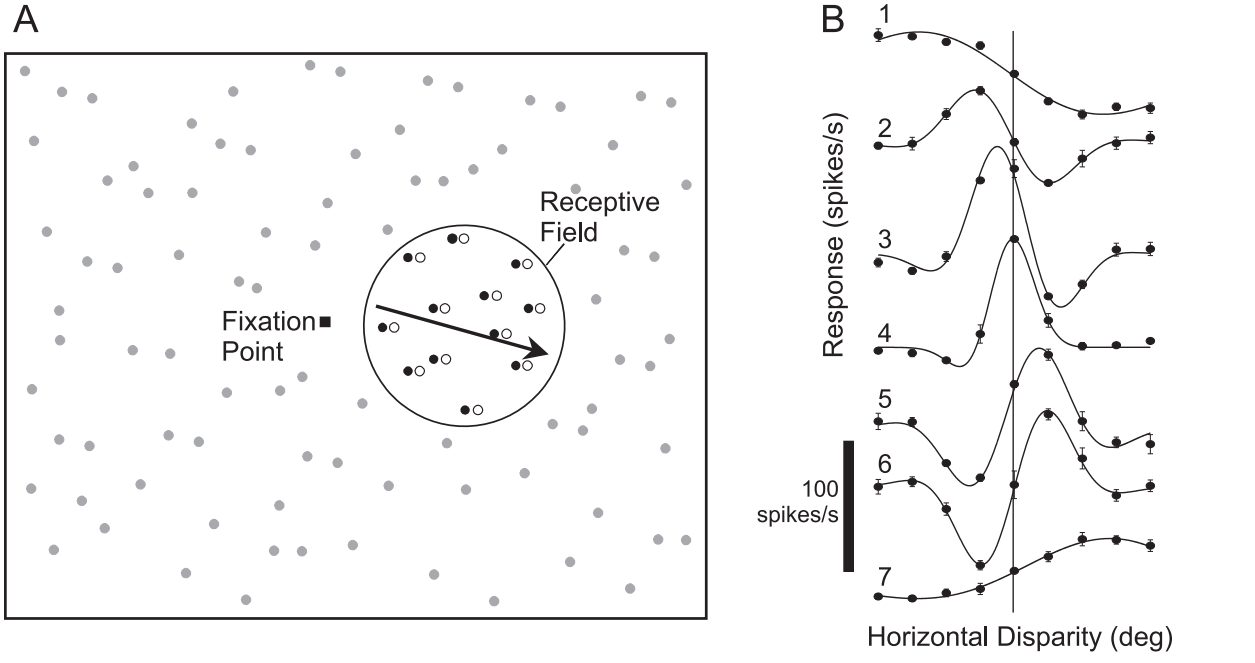


Figure 33.2 Schematic illustration of random-dot stereogram stimulus and example disparity tuning curves measured with this stimulus. (*A*) A circular patch of moving dots having variable disparity was presented over the receptive field (circle, not present in the actual display) of an MT neuron. Solid and open dots within the receptive field denote the images seen by the left and right eyes; the separation between each pair of open and solid dots is the binocular disparity. The remainder of the screen was filled with stationary dots presented with zero disparity (gray dots). The monkey was required to maintain fixation on the fixation point during each trial. (*B*) Disparity tuning curves for seven representative MT neurons. Solid symbols and error bars show the mean response to each disparity ± standard error. The smooth curve through each data set is the best-fitting Gabor function. Neurons are presented (from top to bottom) in order of their preferred disparities, from large Near to large Far. The vertical scale bar corresponds to 100 spikes per second. (Adapted from DeAngelis & Uka, 2003.)

evidence for such a map in area V1 of monkeys (Chen, Lu, & Roe, 2008; Prince et al., 2002). However, there is clear evidence for a map of disparity in both areas V2 (Chen et al., 2008; Nienborg & Cumming, 2006) and V3 (Adams & Zeki, 2001), and it is possible that the map of disparity in MT is at least partially inherited from one or both of these sources.

Absolute and Relative Disparity Selectivity Over the past several years, an important distinction to emerge in the cortical processing of binocular disparity involves the difference between absolute and relative disparity coding. Absolute disparity refers to the interocular difference in the angle subtended by a point in space relative to the projection of the fixation point, which lands on the fovea in each eye (figure 33.3*A*). Thus absolute disparities are defined relative to the point of ocular fixation. In contrast, the relative disparity between two points in space refers to the difference in their absolute disparities (figure 33.3*B*). This distinction becomes especially important when one considers changes in the fixation distance of a subject and hence the vergence angle of the eyes. When we converge or diverge our eyes to focus on a near or far object, respectively, all of the absolute disparities change. On the other hand, the relative disparity between two points in space is unaffected by changes in vergence angle. This means that uncontrolled errors and variations (i.e., noise) in vergence state will affect absolute disparities much more than relative disparities.

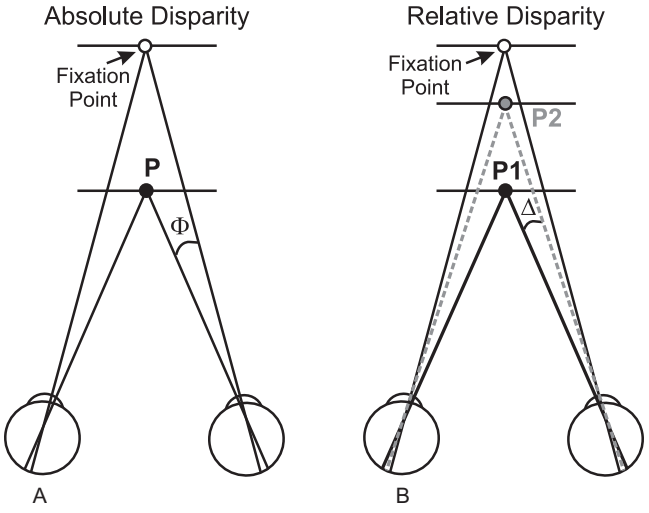


Figure 33.3 Geometric definitions of absolute disparity (**Φ**, left) and relative disparity (Δ, right). Each panel shows a top-down view in which the eyes are converged on a fixation point (open symbol) directly in front of the subject. (*A*) The absolute disparity of point P refers to the angle, **Φ**, subtended by this point relative to the point of fixation. Thus when the subject converges or diverges their eyes to focus at a different distance, the absolute disparity of point P will change. (*B*) The relative disparity of point P1 with respect to point P2 is given by the angle, Δ, which is the difference between the absolute disparities of these two points. If the eyes converge at a different distance, the relative disparity between P1 and P2 will be unchanged.

For this reason, it has been hypothesized that the visual system might need to contain a neural representation of relative disparities to allow precise discrimination of small differences in disparity (Neri, 2005; Neri, Bridge, & Heeger, 2004; Parker, 2007; Prince et al., 2000; Thomas, Cumming, & Parker, 2002). This hypothesis is supported by behavioral evidence. Both humans and monkeys are able to discriminate much smaller differences in the disparity of a target when that target is located close to a reference disparity (Prince et al., 2000; Westheimer, 1979), suggesting that neurons somewhere in the visual system locally compute relative disparities.

To determine whether cortical neurons represent absolute or relative disparities, Thomas and colleagues (2002) devised a test in which two patches of random-dot stereogram are presented in a concentric (center-surround) arrangement and the disparity of both the center and surround are varied in a fully crossed design (figure 33.4*A*). For a neuron that responds solely to the relative disparity between center and surround, the disparity tuning curve in response to the center disparity should shift with the disparity of the surround (figure 33.4*B*). In contrast, no such shift would be observed for a neuron tuned to absolute disparity (figure 33.4*C*). In their ground-breaking study, Thomas and colleagues (2002) found that neurons in area V1 signal absolute disparity, confirming a previous report (Cumming & Parker, 1999), whereas a subset of neurons in area V2 signal relative disparity.

More recently, we have performed similar tests in area MT and have found that MT neurons generally signal absolute disparities, similar to neurons in V1 (Uka & DeAngelis, 2006). This is seen as a distribution of shift ratios (figure 33.4*D*) that cluster around zero, whereas neurons that are tuned for relative disparity would have a shift ratio near one. Fujita and colleagues have recently performed the same test in area V4 and have reported that the majority of V4 neurons have shift ratios greater than zero, with some neurons representing purely relative disparity while many others show an intermediate representation (shift ratios near 0.5) (Umeda, Tanabe, & Fujita, 2007). These comparative results across areas are summarized in figure 33.4*E* (modified from Umeda et al., 2007).

The finding of relative disparity tuning in area V4 but not MT suggests that the ventral processing stream may emphasize relative disparities, whereas the dorsal stream emphasizes absolute disparities. These results are consistent with the results of a functional magnetic resonance imaging (fMRI) study in humans that used an adaptation paradigm to test for sensitivity of different cortical areas to absolute and relative disparities (Neri et al., 2004). This study found that ventral stream areas adapt to both absolute and relative

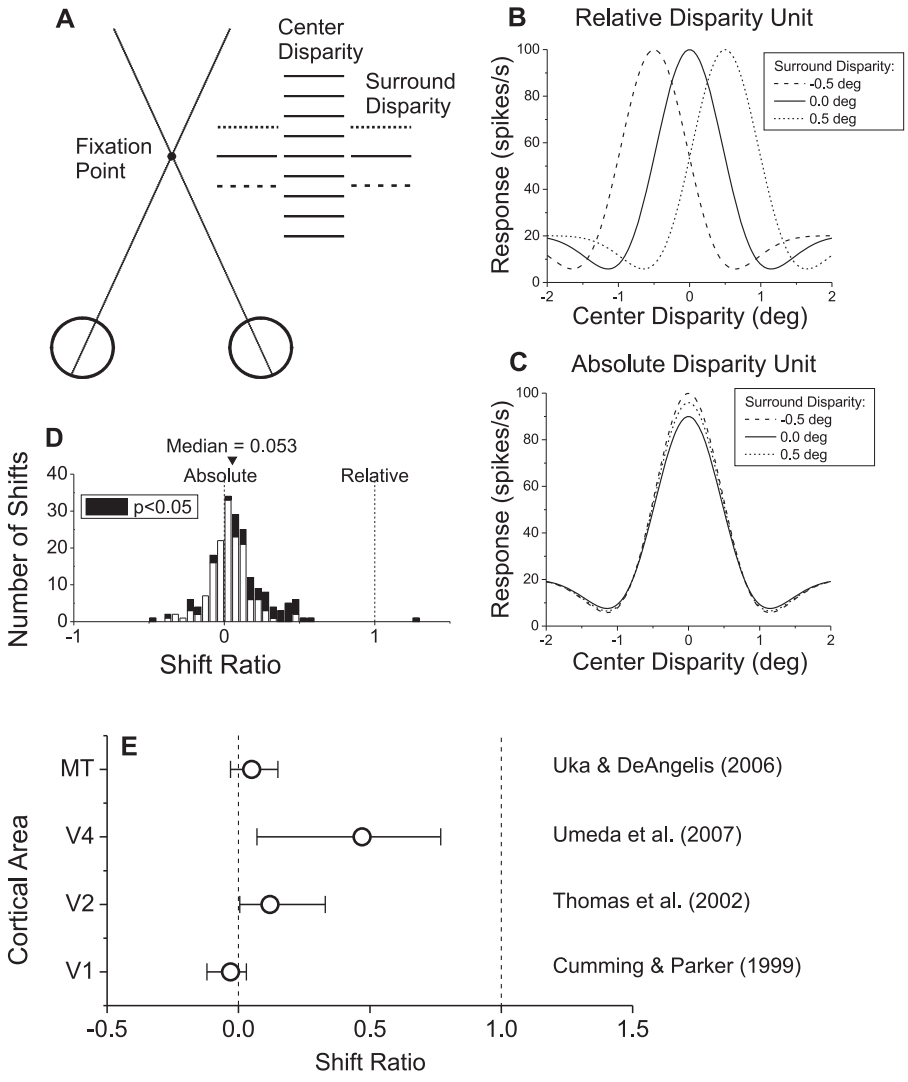


FIGURE 33.4 Stimuli, predicted outcomes, and results of tests for absolute versus relative disparity tuning. (*A*) Top-down view of the stimulus configuration, consisting of a center patch of dots and a surrounding annulus. All combinations of nine center disparities and three to five surround disparities were presented in randomly interleaved trials. (*B*) If a neuron signals relative disparity, the disparity tuning in response to the center patch should shift horizontally by an amount equal to the change in surround disparity. (*C*) If a neuron signals absolute disparity, no shifts should be seen, although some amplitude variations may occur. (*D*) Distribution of shift ratios for 201 pairings of surround disparities from 45 MT neurons. The shift ratio is computed as the horizontal shift of the tuning curve divided by the difference between the two surround disparities. Thus an idealized relative disparity neuron will have a shift ratio of 1, and an idealized absolute disparity neuron will have a shift ratio of 0. Note that shift ratios from area MT are distributed around zero, with a slight but significant bias toward positive values (sign test, $p < 0.0001$). Solid bars denote shift ratios that were significantly different from zero (sequential *F*-test, $p < 0.05$; 52/201 shifts). (Panels *A–D* were adapted from Uka & DeAngelis, 2006.) (*E*) Summary of results of the relative disparity test across studies of four different visual areas, adapted from Umeda and colleagues (2007). For each area, the open symbol indicates the median shift ratio, and the error bars represent the range from the 25th percentile to the 75th percentile. Data were compiled by Umeda and colleagues (2007) across four studies, as indicated. Note that relative disparity selectivity increases from V1 to V2 to V4 (presumably reflecting ascension of the ventral stream), whereas neurons in area MT show absolute disparity tuning.

disparities, whereas dorsal stream areas adapt only to absolute disparities. Together, these findings from monkeys and humans are consistent with the hypothesis that the ventral stream carries out sophisticated disparity computations to represent the 3D shape of objects, whereas the dorsal stream carries out somewhat simpler computations that are aimed at computing the location of objects in 3D space and also perhaps at representing the coarse layout of surfaces in the scene (see also Neri, 2005; Parker, 2007). Along these lines, we might expect area MT to contribute to coarse judgments of depth based on absolute disparities but not to fine judgments of depth based on relative disparities. We shall return to this prediction later.

Coding of Three-Dimensional Surface Orientation Most physiological studies of disparity processing in visual cortex have examined responses to frontoparallel planar surfaces that vary in distance from the observer. In natural scenes, however, binocular disparity varies smoothly across surfaces that can have many possible 3D orientations relative to the observer. Spatial gradients of disparity thus provide powerful cues to 3D object shape and 3D surface orientation (Howard & Rogers, 1995, 2002). The simplest form of spatial variation involves monotonic gradients of disparity across space that specify the 3D orientation of planar surfaces. As illustrated in figure 33.5*A*, the 3D orientation of planar surfaces can be parameterized in terms of *tilt* and *slant*.

Whereas human perception of tilt and slant defined by gradients of disparity has been studied considerably (Howard & Rogers, 2002; Sedgwick, 1986), only recently have physiologists examined how cortical neurons represent 3D orientation. In area MT, we have tested neurons with random-dot stereograms containing linear gradients of disparity, and we have found that more than half of MT neurons show significant tuning for the tilt of planar surfaces (Nguyenkim & DeAngelis, 2003). This property is shown for an example neuron in figure 33.5*C*. The neuron shows broad but robust tuning for the tilt of the surface, and this tuning is maintained across variations in the mean disparity of the stimulus. This insensitivity to variations in mean disparity around the peak of the cell's disparity-tuning curve (figure 33.5*B*) indicates that tilt selectivity cannot be simply explained by miscentering the gradient stimulus on the receptive field or by receptive field inhomogeneities (Nguyenkim & DeAngelis, 2003).

Interestingly, this finding of invariant tilt tuning suggests that MT neurons possess some form of relative disparity selectivity when tested with disparity gradients, whereas they do not show this property when tested with concentric edges (figure 33.4*D*). Although the mechanisms underlying this difference remain unclear, this comparison highlights the important point that there is no single unique test for relative

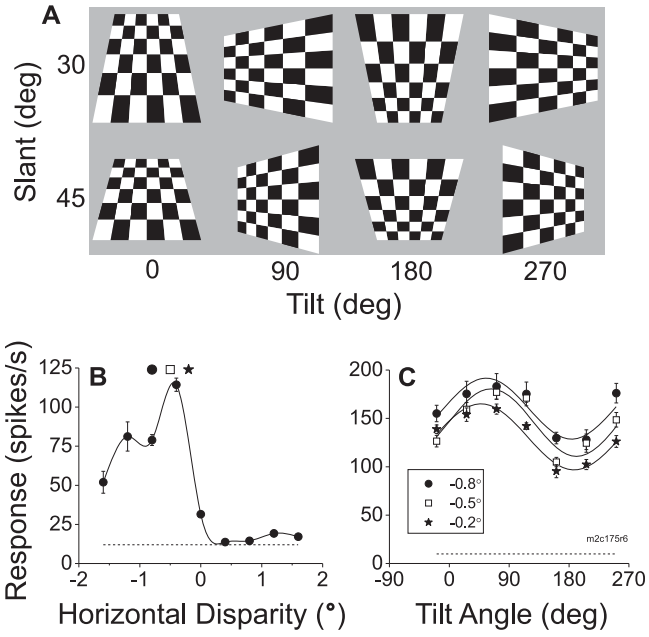


Figure 33.5 Schematic illustration of the 3D orientation of planar surfaces, parameterized by tilt and slant, as well as data from a tilt-selective MT neuron. (*A*) *Tilt* refers to the axis around which the plane is rotated away from frontoparallel, and *slant* defines the amount by which the plane is rotated. Zero slant corresponds to a frontoparallel surface for which the tilt is undefined. In this illustration, tilt and slant are defined by perspective and texture gradient cues. In the MT experiments, surface orientation was defined solely by the direction and magnitude of a linear gradient of horizontal disparity in a random-dot stereogram. (*B*) A conventional disparity-tuning curve for an MT neuron measured using random-dot stereograms (slant was zero, and different uniform horizontal disparities were applied). Mean responses ± standard error are shown for each stimulus disparity, along with a spline fit. Symbols at the top indicate the three mean disparities used for testing tilt selectivity with disparity gradients. (*C*) Tilt-tuning curves for the same MT neuron are show at three different mean disparities (coded by symbol shape). Smooth curves indicate the best fits of the modified sinusoid function. Note that the neuron shows clear tuning for the tilt of a planar random-dot surface and that this tilt tuning is robust to changes in the mean disparity of the stimulus. (Adapted from Nguyenkim & DeAngelis, 2003.)

disparity selectivity, and it might depend greatly on stimulus geometry. In MT, tilt tuning is seen only for stimuli that are presented at large slants (generally >45 degrees) (Nguyenkim & DeAngelis, 2003), suggesting that 3D orientation coding in MT involves a rather coarse mechanism that is more likely to be involved in providing the basic layout of surfaces in the scene rather than in analyzing the details of 3D shape (see also Parker, 2007), though this remains to be tested further.

Selectivity for 3D surface orientation has also been observed in other cortical areas. In the caudal intraparietal area (CIP), Sakata and colleagues have described neurons that signal the tilt of surfaces defined by disparity gradients

(Taira et al., 2000; Tsutsui et al., 2001, 2002). Hinkle and Connor (2002) have also described neurons in area V4 that are selective for the 3D orientation of bar stimuli, although this selectivity could be driven more by orientation differences between the eyes because the same neurons did not generally show tuning for tilt and slant in disparity-defined random-dot surfaces. In the inferotemporal (IT) cortex, Janssen, Orban, and colleagues have conducted an impressive series of studies showing that IT neurons have selectivity for 3D shape defined by gradients of disparity as well as boundary cues (Janssen, Vogels, Liu, & Orban, 2001; Janssen et al., 1999, 2000a, 2000b). Thus it is clear that neurons at the upper levels of both the dorsal and ventral processing streams make use of spatial gradients of disparity to extract information about both 3D shape and object/surface orientation. However, much remains to be learned, and the respective roles of these areas in perception of 3D structure are still not well understood.

In addition to disparity gradients, the 3D orientation of surfaces may be specified by gradients of texture, velocity, or luminance (shading) (Sedgwick, 1986). Neurons that are closely involved in perception of 3D orientation may thus be expected to signal tilt and slant based on multiple cues. In area CIP, Sakata's group has shown that neurons signal tilt by gradients in both disparity and texture (presented separately), and that their tilt preferences for the two cues are often well matched (Tsutsui et al., 2002). However, no published study has examined how neurons respond to multiple cues to 3D orientation presented simultaneously. In area MT, neurons have previously been shown to exhibit selectivity for tilt defined by velocity gradients (Treue & Andersen, 1996; Xiao, Marcar, Raiguel, & Orban, 1997). We have presented preliminary evidence that individual MT neurons are tuned for tilt defined by both disparity and velocity gradients (Nguyenkim & DeAngelis, 2004). Some neurons have matched tilt tuning for the two cues, whereas others do not. Responses to both cues together appear to be well predicted by weighted linear summation of the individual cue responses (unpublished). It is currently unclear whether MT plays a role in perceptual cue integration for 3D orientation perception or whether it may simply be an early stage at which disparity and velocity gradients begin to interact. A recent fMRI study, performed by using behaving macaques, suggests that there is considerable additional processing of 3D surface orientation and shape in regions of the intraparietal sulcus that receive inputs from area MT (Durand et al., 2007).

Linking neural representation to function: Roles of area MT in coarse and fine depth discrimination

As was discussed above, binocular disparity information is now known to be represented across a broad range of visual cortical areas in primates. There seem to be two main possibilities for why this might occur: (1) Disparity processing is highly distributed such that most aspects of depth perception depend on simultaneous activation of many regions of cortex, or (2) different cortical areas have specialized representations of binocular disparity that are well suited to some tasks but not others. In the latter scenario, depth perception in a specific context could depend on only a small subset of visual areas or perhaps only on a subset of neurons within a single visual area.

Thus far, we have already discussed evidence that favors the notion of specialized representations, namely, that ventral stream areas appear to represent the precise relative disparity information that is thought to be needed for fine depth discrimination and 3D shape perception, whereas the dorsal stream appears to emphasize absolute disparities (Thomas et al., 2002; Uka & DeAngelis, 2006; Umeda et al., 2007). Moreover, it is likely that we know about only a small fraction of the differences between areas and between visual streams at this time. If different cortical areas are specialized to perform different tasks, then it should be possible to identify experimentally the areas and/or neurons that contribute to performance of a particular task. In recent years, my laboratory has attempted to clarify the functional roles of area MT in stereo vision by performing a series of experiments with monkeys that were trained to perform tasks that were chosen to reveal differences in function that may be linked to absolute versus relative disparity representations.

Coarse and Fine Depth Discrimination Tasks To probe depth perception based on coarse absolute disparity information, we trained monkeys to perform the Coarse task illustrated in figure 33.6*A*. In this task, dots in a stereogram are divided into two groups with adjustable percentages: "signal" dots are all presented at the same disparity in each trial, which is near or far relative to the plane of fixation; "noise" dots are given random disparities in each trial such that they form a 3D cloud. The monkey's task is to judge whether the net depth of the stimulus is near or far and to make a saccadic eye movement to signal its choice (Uka & DeAngelis, 2003). Across trials, the relative proportion of signal and noise dots, indexed by a variable called *binocular correlation*, is varied to manipulate task difficulty. Figure 33.7*A* (open symbols) shows a psychometric function for one monkey in a typical session. Note that in this task, the monkeys always discriminated between two signal disparities (e.g., −0.4° and +0.4°) that were on opposite sides of zero disparity and were well above stereoacuity thresholds.

To assay the contribution of neurons to depth perception based on fine relative disparities, we also trained monkeys to perform the Fine task depicted in figure 33.6*B*. In this task, a bipartite center-surround random-dot stereogram is presented, and the monkey is required to report whether

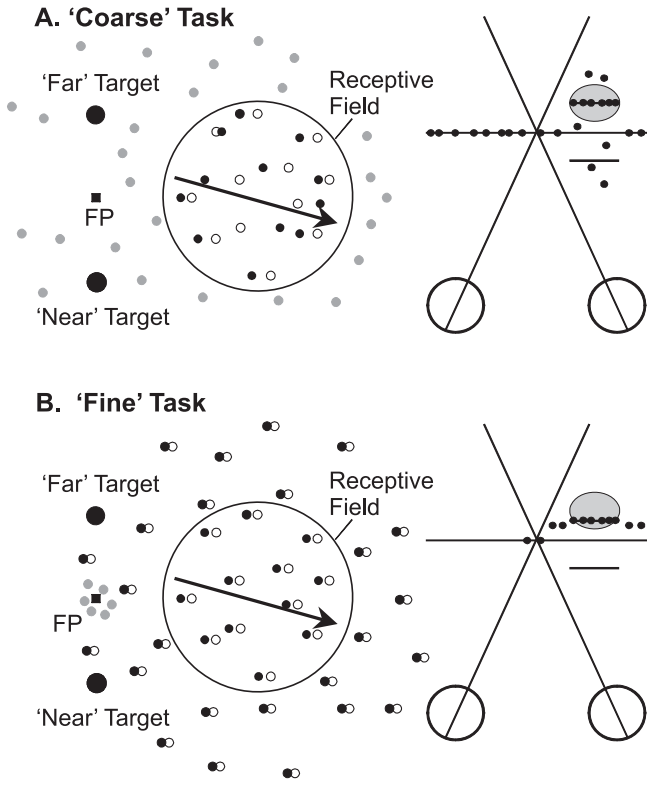


FIGURE 33.6 Schematic illustration of two depth discrimination tasks used to study functional contributions of area MT. (*A*) The Coarse task. A random-dot stereogram was presented over the receptive field (RF), and dots moved at the neuron's preferred velocity (arrow). Solid and open dots represent left and right half-images, respectively. The background was covered with dynamic zero-disparity dots (gray). Saccade targets were located 5° above and below the fixation point, corresponding to far and near choices, respectively. The strength of the depth signal was adjusted by varying binocular correlation. At 50% binocular correlation (right), half of the dots within the receptive field were presented at either the neuron's preferred disparity (horizontal line inside gray oval) or the disparity that elicited a minimal response (null disparity). The remaining dots had random disparities. (*B*) The Fine task. A bipartite (center-surround) random-dot stereogram was presented. The center patch covered the RF and contained dots moving at the preferred velocity (arrow). The surrounding annulus contained stationary dots presented (in most cases) at a nonzero disparity. A small patch of zero-disparity dots (gray) surrounded the fixation point to help anchor vergence. The monkey reported whether the center patch was in front of or behind the surround patch, and task difficulty was manipulated by finely varying the center disparity around the surround disparity. (Adapted from Uka & DeAngelis, 2006.)

the center patch of dots appears near or far relative to the surround (Uka & DeAngelis, 2006). Both center and surround are presented without noise (100% binocular correlation), and the relative disparity between center and surround is varied in fine steps to measure psychophysical threshold (e.g., figure 33.7*B*). Importantly, the absolute disparities of the center and surround could be both far or both near, such that monkeys are required to judge relative depth to achieve high performance on this task.

Given that neurons in area MT have fairly broad disparity tuning and do not represent relative disparities in a center-surround configuration (figure 33.4*D*), we hypothesized that MT would play a significant role in the Coarse task but not the Fine task. This was assessed by using a variety of approaches, as described below (see also Parker & Newsome, 1998).

NEURONAL VERSUS BEHAVIORAL SENSITIVITY From the distributions of firing rates measured during performance of the tasks, we used ROC analysis to compute the ability of an ideal observer to discriminate depth on the basis of the responses of each single MT neuron (Uka & DeAngelis, 2003, 2006). Example neurometric functions for representative MT neurons are shown in figure 33.7*A* for the Coarse task and figure 33.7*B* for the Fine task (solid symbols). These neurometric functions describe how the performance of the ideal observer increases as the differences between near and far stimuli become more salient. From each such data set, we computed psychophysical and neuronal thresholds as the stimulus values at which performance reaches 82% correct. Thus each experiment yielded both a psychophysical and a neuronal threshold that could be compared quantitatively.

Figure 33.7*C* shows the distribution of the ratio of neuronal-to-psychophysical thresholds for 104 MT neurons studied during the Coarse task (Uka & DeAngelis, 2003). While threshold ratios span a wide range, the average ratio (geometric mean = 0.98) was close to unity, indicating that the average MT neuron could discriminate coarse disparities in noise with sensitivity comparable to that of the animal. This result is very similar to that found by Newsome and colleagues for direction discrimination in MT (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1992). Thus neuronal sensitivity suggests that area MT could account for coarse depth discrimination.

Figure 33.7*D* shows the analogous distribution of threshold ratios for 98 neurons that were tested during the Fine task. In this case, the average threshold ratio (1.76) is closer to 2, indicating that MT neurons are not as sensitive as the animal is. However, the best neurons could be sufficiently sensitive to account for behavior.

CHOICE PROBABILITIES In a psychophysical task performed around threshold, the same (weak) stimulus gives rise to different perceptual reports, as well as different neural responses, across repeated trials. By testing for a correlation between the trial-to-trial fluctuations in perceptual reports and neural responses (choice probability, or CP), it may be possible to identify neurons that are functionally coupled to perceptual decisions (Britten et al., 1996; Krug, 2004). An advantage of this approach is that it affords single-cell

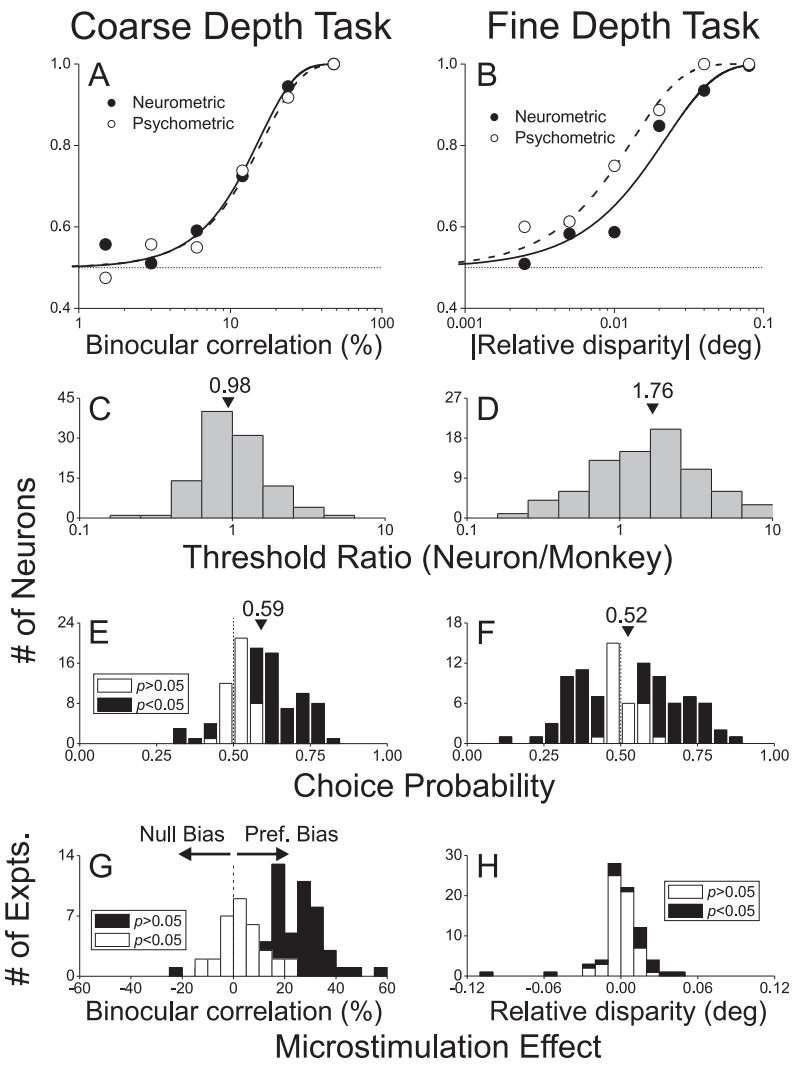


FIGURE 33.7 Summary of single-unit and microstimulation experiments performed in area MT using both the Coarse (left column) and Fine (right column) depth tasks. (*A*) Example data from a typical experiment using the Coarse task. The psychometric function (open symbols) shows the monkey's percentage of correct responses as a function of binocular correlation. The neurometric function (solid symbols) shows the predicted performance of an ideal observer based on the responses of a single MT neuron. In this example, the neuron has sensitivity nearly identical to that of the animal. (*B*) Example psychometric and neurometric functions for a typical experiment using the Fine task. In this case, performance is plotted as a function of the (unsigned) relative disparity between center and surround stimuli. In this experiment, the neuron is about half as sensitive as the monkey. (*C*) Distribution of the ratio of neuronal/psychophysical threshold ratios across 104 recording sessions involving the Coarse task. The geometric mean ratio was 0.98 (arrowhead). (*D*) Distribution of neuronal/psychophysical threshold ratios for the Fine task (N = 98). The geometric mean was 1.76. (*E*) Distribution of choice probabilities for the Coarse task (N = 104). CP values significantly different from 0.5 are indicated by solid bars. Note that most CP values are greater than 0.5. (*F*) Distribution of choice probabilities for the Fine task (N = 98). (*G*) Distribution of microstimulation effects for the Coarse task (N = 78). Solid bars denote individually significant effects. Positive values indicate biases toward the preferred disparity of the stimulated neurons, as measured in units of percent binocular correlation for the Coarse task. Most experiments produced a significant preferred bias. (*H*) Distribution of microstimulation effects for the Fine task (N = 46). Shifts are now measured in degrees of relative disparity. Most experiments produced no effect, and the median shift was not significantly different from zero. (Panels *A* and *C* were adapted from Uka & DeAngelis, 2003; panel *E* was adapted from Uka & DeAngelis, 2004; panel *G* was adapted from DeAngelis et al., 1998, including additional data; panels *D* and *H* were adapted from Uka & DeAngelis, 2006. Data in panel *F* are previously unpublished.)

resolution and allows one to relate the tuning properties of neurons to behavior. Although CPs simply reflect a correlation between neurons and perceptual decisions, there is evidence to suggest that significant CPs reflect a functional contribution of neurons. For example, CPs in area MT and V2 have been shown to vary according to the tuning properties of neurons in a manner that appears to reflect the animal's (suboptimal) strategy for solving the task (Nienborg & Cumming, 2007; Uka & DeAngelis, 2004).

We have used CPs as another means to evaluate the role of area MT in the Coarse and Fine depth tasks. As is shown in figure 33.7*E*, most MT neurons have CPs greater than 0.5 in the Coarse task, which indicates that the neurons tend to fire more strongly when the monkey reports that an ambiguous (e.g., 0% binocular correlation) stimulus matches the preferred depth of the neuron (Uka & DeAngelis, 2004). The average CP is 0.59, which is significantly greater than 0.5 ($p \ll 0.001$). This means that an ideal observer could predict the choices of a monkey with 59% correct accuracy by monitoring the activity of an average MT neuron. This finding is consistent with the notion that area MT makes an important contribution to the Coarse depth task.

Recently, Nienborg and Cumming (2006, 2007) have examined responses of neurons in area V1 and V2 during performance of a task that is nearly identical to our Coarse task. Interestingly, neurons in V2 show CPs comparable to those we have seen in MT, whereas neurons in V1 do not. Thus trial-to-trial variability in the representation of disparities in V1 does not seem to be linked to depth percepts, whereas similar variability in V2 and MT does correlate with percepts. It is currently unclear whether neurons in V2 with significant CPs in the Coarse task reside in the portions of V2 (the thick stripes) that project heavily to area MT. However, a recent study shows that inactivating areas V2 and V3 by cooling substantially reduces disparity selectivity in MT (Ponce, Lomber, & Born, 2008), consistent with the idea that MT may inherit at least some of its disparity selectivity from V2.

It is also interesting to note that CPs for the Coarse task have thus far been seen in areas (V2, MT) that contain a topographic representation of disparity (Chen et al., 2008; DeAngelis & Newsome, 1999) but not in area V1, which lacks such a map for disparity (Chen et al., 2008; Prince et al., 2002). It will be fascinating to see whether a correlation between functional architecture and CPs emerges as similar data are collected from additional tasks in additional areas.

Figure 33.7*F* shows analogous CP data from area MT during performance of the Fine task (unpublished data). In this case, the mean CP is 0.52, which is not significantly different from 0.5 ($p > 0.05$). This may be taken as evidence that area MT does not contribute to performance of the Fine task. Note, however, that many more neurons than expected by chance have CPs significantly different from 0.5 (solid bars). These neurons are significantly correlated with perceptual decisions about fine relative disparities, but there is no consistent relationship between firing rates and choices across the population. Thus, on average, one cannot reliably predict a monkey's choices by measuring the response of MT neurons in the Fine task. However, this pattern of results remains somewhat puzzling, as it is not clear why MT neurons should show significant CPs that are both lower and higher than 0.5. If one speculates that CPs arise through a top-down signal from decision circuitry to MT, then it might be that these top-down signals cannot correctly target MT neurons because MT does not contain a topographically organized representation of relative disparities.

Causal Manipulations Choice probabilities establish a correlation between neural responses and perceptual decisions (independent of the physical stimulus) but do not establish a causal contribution of those neurons to perception. To further link a particular visual area with specific functions in 3D vision, we need to directly manipulate neural activity during performance of relevant tasks. One approach involves electrical microstimulation (Cohen & Newsome, 2004; Salzman, Murasugi, Britten, & Newsome, 1992; Tehovnik, 1996), in which weak biphasic current is passed through a recording electrode to activate a cluster of neurons near the tip of the electrode whose tuning properties are known. By placing an electrode into the midst of one of the disparity columns in area MT (DeAngelis & Newsome, 1999), we have used microstimulation to probe the causal contribution of area MT to the Coarse and Fine tasks.

In the Coarse task, microstimulation systematically biases monkeys' judgments of depth, as summarized in figure 33.7*G* (see also DeAngelis, Cumming, & Newsome, 1998; Uka & DeAngelis, 2006). A positive effect of microstimulation means, for example, that electrical stimulation of a cluster of near-tuned neurons causes the monkey to report stimuli as "near" significantly more often than occurs when microstimulation is withheld. Note that microstimulation frequently produced statistically significant effects in the Coarse task (solid bars in figure 33.7*G*) and that the vast majority of these effects were positive. Only one experiment produced a microstimulation effect in the "wrong" direction, such that stimulation of a cluster of far-preferring neurons produced a bias in favor of perceiving near. This finding, coupled with the sensitivity and CP analyses described above, establishes that area MT contributes to coarse depth perception based on absolute disparities.

Figure 33.7*H* shows comparable microstimulation results for the Fine task. In this case, the median effect of microstimulation is not significantly different from zero ($p = 0.88$). Moreover, when microstimulation did produce a significant

effect (solid bars) it was frequently in the wrong direction (Uka & DeAngelis, 2006). Thus we found no clear evidence that area MT contributes to performance of the Fine task, consistent with the hypothesis that MT's role in this task is limited because it does not carry a representation of fine relative disparities.

Together, these findings establish a satisfying connection between the neural representation of disparities in area MT (absolute, not relative) and the functional contributions of area MT to depth perception. These findings are consistent with the idea that different visual areas contain specialized representations of binocular disparity signals, and specifically support the notion that the dorsal stream (including MT) mainly processes absolute disparity information to localize objects in 3D space, whereas the ventral stream emphasizes computations of relative disparity for the purpose of 3D shape perception (Neri, 2005; Parker, 2007). Our findings spur hope that similar studies, employing a variety of tasks in a variety of areas, will be capable of revealing a functional taxonomy of visual cortical areas with respect to their roles in 3D vision. This remains to be seen, but there is reason to be optimistic that we can understand the selective contributions of individual areas to this overall process.

A weakness of the comparisons that we have made between the Coarse and Fine tasks is that several aspects of the stimuli differ between the two tasks, including absolute versus relative disparities, the range of disparities, inclusion of noise, and presence of segmentation boundaries. Going forward, it will be important to design stimuli that allow us to test perception of depth versus 3D shape while eliminating or minimizing differences in other stimulus variables.

Selectivity for depth from motion parallax in area MT

As is illustrated in figure 33.1*B*, self-movement (e.g., moving one's head from side to side) generally causes the image of an object to move on the retina, and both the direction and speed of image motion depend on the location of the object in depth. This depth-dependent image motion is called *motion parallax*. Motion parallax can also arise because of the movement of objects that have depth structure, but here I shall focus on motion parallax resulting from observer movement.

In a ground-breaking series of psychophysical studies, Rogers and Graham placed subjects in an apparatus with a sliding chin rest and asked them to move their heads back and forth while fixating a point on a video display with one eye (Graham & Rogers, 1982; Rogers & Graham, 1979, 1982; Rogers, 1993; Rogers & Rogers, 1992). As the subjects moved their heads (and correspondingly their eyes), the experimenters updated the positions of dots in the display such that their motion was consistent with the presence of a corrugated surface in depth. Despite the lack of disparity or any pictorial depth cues, this arrangement produces a compelling sensation of depth. Studies have shown that depth perception from motion parallax is almost as precise as from disparity (Rogers & Graham, 1982). Moreover, psychophysical studies suggest that disparity and motion parallax processing may share a common neural substrate. Depth percepts from disparity and motion parallax can cross-adapt each other, and combining the cues together can yield substantial improvements in sensitivity over either cue alone (Bradshaw & Rogers, 1996).

Whereas perception of depth from motion parallax has been well studied (Nawrot & Joyce, 2006), the neural basis for this behavior has remained unknown. Surely, a neural substrate for depth from motion parallax requires neurons that are selective for the direction and speed of visual motion, and such neurons can be found in many visual areas in primates beginning with V1. It has also been suggested that neurons with relative motion selectivity (Cao & Schiller, 2003; Li, Lei, & Yao, 1999) might provide important inputs to such a depth mechanism. However, the presence of visual motion selectivity alone, even relative motion selectivity, does not establish that neurons can provide depth information based on motion parallax.

How, then, can we identify neurons that participate in computing depth from motion parallax? Our approach has been to exploit the fact that visual image motion itself can be depth-sign ambiguous, as illustrated in figures 33.8*A* and 33.8*B*. In the absence of pictorial depth cues such as occlusion and size, the retinal image motion generated by near and far objects (having equivalent disparities) is identical except for the phase of the motion relative to movement of the subject's head (or eyes). Objects that are nearer than the point of fixation will move in the direction opposite to head motion, whereas far objects will move in the same direction as the head. Thus in the absence of pictorial cues, neurons must combine retinal image motion with extraretinal signals related to head and/or eye movement to determine the sign of depth from motion parallax. We designed random-dot stimuli that were depth-sign ambiguous by removing all pictorial cues to depth (stimulus size, dot size and density, occlusion, etc.) (Nadler, Angelaki, & DeAngelis, 2008). If neurons simply respond to visual image motion, then they cannot differentiate between our near and far stimuli. On the other hand, if neurons receive extraretinal inputs that specify the phase of visual image motion relative to head or eye motion, then they may become selective for depth sign.

Figures 33.8*C* and 33.8*D* illustrate responses from a neuron recorded from area MT under monocular viewing conditions (ipsilateral eye occluded). In the Retinal Motion condition (figure 33.8*C*), the visual stimulus simulated a

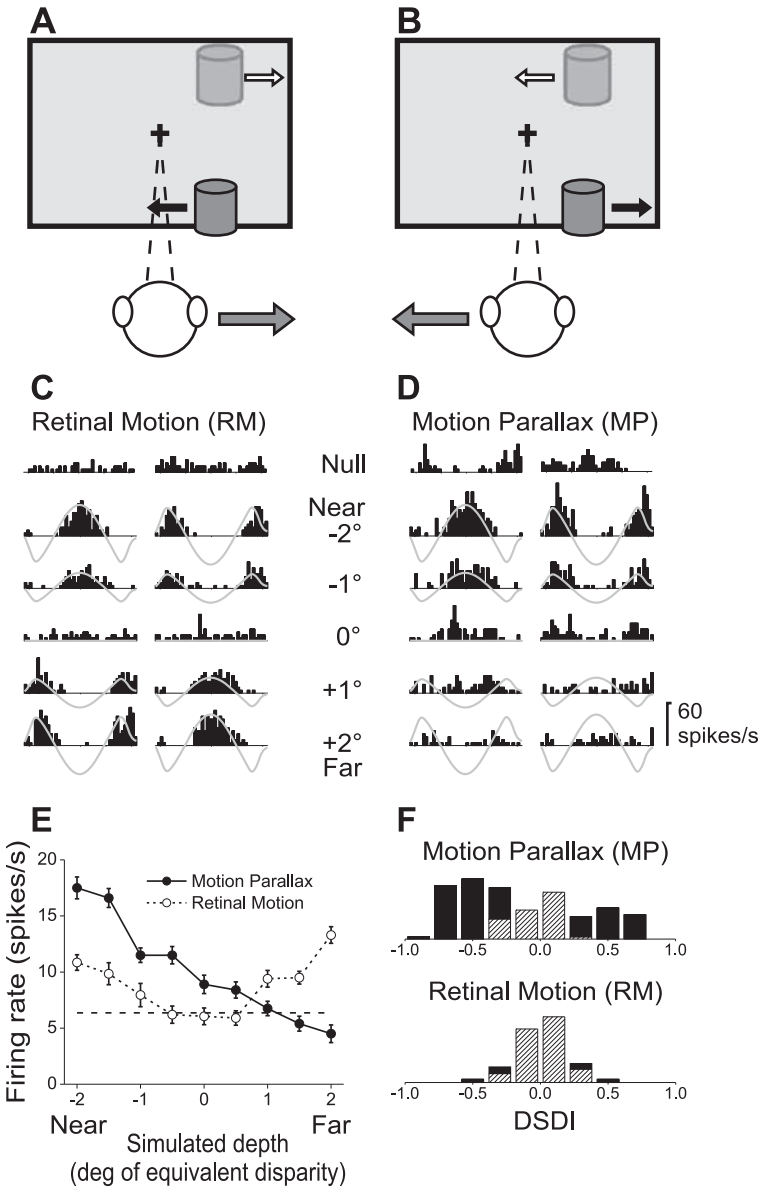


FIGURE 33.8 Motion parallax, depth-sign ambiguity, and a potential solution in area MT. (*A*, *B*) Schematic illustration of how head translation generates motion parallax. As the head moves to the right (*A*), the image of a near object (bottom cylinder) moves to the left while the image of a far object (top cylinder) moves to the right. The opposite occurs during head movement to the left (*B*). If pictorial cues to depth, such as size and occlusion, are not present, the only difference between the near and far objects is how the phase of their visual motion relates to that of the observer's motion. Thus an extraretinal signal related to head movement or eye movement is needed to determine depth sign from motion parallax. (*C*, *D*) Responses of an example neuron showing selectivity for depth from motion parallax. Panel *C* shows PSTHs in response five simulated depths during the Retinal Motion (RM) condition, in which the monkey remains stationary while visual motion simulates depth. One column of poststimulus time histograms (PSTHs) is shown for each starting phase of simulated observer motion. Traces of the retinal velocity of the stimulus within the receptive field are superimposed in gray, with peaks representing image motion in the preferred direction and troughs indicating motion in the anti-preferred direction. The depth-sign ambiguity in the retinal motion stimulus is reflected in the responses of this MT neuron, which are symmetric in strength about 0° in the RM condition. PSTHs labeled "Null" represent responses obtained when no random dots were presented in the receptive field. (*D*) Responses of the same neuron in the Motion Parallax (MP) condition. Note that responses to near stimuli are accentuated, while responses to far stimuli are suppressed relative to the RM condition. A modest modulation can be seen in the "Null" condition, reflecting a response associated with head/eye movement. (*E*) Depth-tuning curves for the same neuron are shown for the RM (open symbols) and MP (solid symbols) conditions. This neuron prefers near stimuli in the MP condition. DSDI values are −0.73 and 0.15 in the MP and RM conditions, respectively. Error bars are standard errors. The dashed horizontal line represents spontaneous activity. (*F*) Distributions of the depth sign discrimination index (DSDI) are shown for 144 MT neurons tested during the MP (top) and RM (bottom) conditions. DSDI values significantly different from zero are indicated by solid bars. (Adapted from Nadler et al., 2008.)

surface placed at different distances from the animal, but the monkey's head and eyes remained still such that the stimulus was depth-sign ambiguous. In this case, the neuron responds more strongly as stimuli lie farther away from the plane of fixation (which has an equivalent disparity of 0°), but the response is very similar for stimuli of opposite depth signs (e.g., −1° versus +1°). This can be seen further when the data are plotted as a tuning curve (figure 33.8*E*, open symbols). The tuning curve for the Retinal Motion condition is symmetric around 0°, indicating that the neuron does not distinguish the sign of depth. In the Motion Parallax condition (figure 33.8*D*), the trained animal is translated side-to-side by a motion platform while maintaining ocular fixation on a world-fixed target. Thus as the animal is translated rightward, it makes a smooth leftward eye movement to keep focused on the fixation target. The retinal image motion in this condition is the same as that generated (artificially) in the Retinal Motion condition, but now the monkey's movement provides extraretinal signals. Under this condition, the same neuron now responds more strongly to near stimuli and more weakly to far stimuli (figure 33.8*D*). Plotted as a tuning curve (figure 33.8*E*, filled symbols), the responses show a clear preference for near stimuli over far stimuli in the Motion Parallax condition. Since retinal image motion was the same in these two stimulus conditions (see Nadler et al., 2008, for details), this difference in tuning must arise from the interaction between extraretinal signals and visual motion information.

Figure 33.8*F* summarizes results from a population of 144 MT neurons, as quantified by a depth sign discrimination index (DSDI), which takes on values near −1 for strong near tuning, values near 0 for no depth-sign tuning, and values near +1 for strong far tuning. In the Motion Parallax condition, DSDI values take on a broad range of values (figure 33.8*F*, top), and ~70% of MT neurons show significant selectivity for depth sign based on motion parallax (filled bars). In the Retinal Motion condition, the distribution of DSDI is more tightly centered around zero with few cells passing the test for significance (figure 33.8*F*, bottom). Thus a majority of neurons in area MT are able to combine visual image motion with extraretinal signals to compute the sign of depth based on motion parallax.

These findings provide the first clear evidence for a population of neurons that represent depth from motion parallax and establish a second major neural mechanism of depth perception in the brain. These results also raise many questions for additional research:

1. Is selectivity for depth-sign created de novo in area MT, or does it occur at earlier stages in the visual pathways? This can be tested by performing similar experiments in areas that provide substantial inputs to MT, namely, V1, V2, and V3.

2. Are the depth-sign signals in area MT actually used by monkeys to judge depth from motion parallax? Alternatively, these findings could be a by-product of interactions between visual and extraretinal signals for some other purpose. This can be tested by training monkeys to discriminate depth sign from motion parallax and by applying electrical microstimulation to MT.

3. What is the extraretinal signal that generates depth-sign selectivity in MT? By the design of our Motion Parallax condition, it could be either a vestibular (or proprioceptive) signal related to head movement or a smooth eye movement signal. Our preliminary results (Nadler, Angelaki, & DeAngelis, 2006) indicate that the extraretinal signal is a smooth eye movement command, as has been suggested by human psychophysical studies (Naji & Freeman, 2004; Nawrot, 2003a, 2003b; Nawrot & Joyce, 2006).

4. How is the depth tuning of MT neurons for motion parallax related to their selectivity for binocular disparity? Our preliminary results (unpublished) show that some MT neurons have matched depth-sign preferences for disparity and motion parallax (e.g., both near or both far), whereas other neurons curiously have opposite depth-sign preferences for the two cues. The former neurons could play a role in integrating the two depth cues to achieve greater perceptual sensitivity, as demonstrated psychophysically (Bradshaw & Rogers, 1996). The role of the latter neurons with opposite preferences is unclear, but they might play a role in distinguishing retinal image motion that arises because of the depth structure of the scene from image motion that arises due to the motion of objects in the scene. Area MT should provide an excellent model system for studying how depth cues are integrated both perceptually and neurally.

Conclusion

The past decade has seen a dramatic increase in our knowledge of the neural basis of depth perception. New cortical areas have been found to represent binocular disparities, new representations of both disparity and motion parallax information have been uncovered, the first topographic maps of binocular disparity have been measured, and the first causal links between neural activity and depth perception have been established. During this period, the collective efforts of several laboratories have established 3D vision as a highly productive model system for understanding the neural basis of perception. Here I have summarized some of the work from area MT that has contributed to our current knowledge and that helps to provide a roadmap for future studies. One of the great remaining challenges, which is applicable to almost any sensory or motor system, is to understand how neural representations are specialized in different areas and how these specialized representations guide behavior.

ACKNOWLEDGMENTS I thank several former members of my laboratory who have conducted the research described here, including Takanori Uka, Ben Palanca, Jerry Nguyenkim, and Jacob Nadler. I also thank Dora Angelaki, who has been a valuable collaborator on some of this work.

REFERENCES

ADAMS, D. L., & ZEKI, S. (2001). Functional organization of macaque V3 for stereoscopic depth. *J. Neurophysiol.*, *86*, 2195–2203.

ALBRIGHT, T. D., DESIMONE, R., & GROSS, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *J. Neurophysiol.*, *51*, 16–31.

BARLOW, H. B., BLAKEMORE, C., & PETTIGREW, J. D. (1967). The neural mechanism of binocular depth discrimination. *J. Physiol. Lond.*, *193*, 327–342.

BORN, R. T., & BRADLEY, D. C. (2005). Structure and function of visual area MT. *Annu. Rev. Neurosci.*, *28*, 157–189.

BRADSHAW, M. F., & ROGERS, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vis. Res.*, *36*, 3457–3468.

BRITTEN, K. H., NEWSOME, W. T., SHADLEN, M. N., CELEBRINI, S., & MOVSHON, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neurosci.*, *13*, 87–100.

BRITTEN, K. H., SHADLEN, M. N., NEWSOME, W. T., & MOVSHON, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J. Neurosci.*, *12*, 4745–4765.

CAO, A., & SCHILLER, P. H. (2003). Neural responses to relative speed in the primary visual cortex of rhesus monkey. *Vis. Neurosci.*, *20*, 77–84.

CHEN, G., LU, H. D., & ROE, A. W. (2008). A map of horizontal disparity in primate V2. *Neuron*, *58*, 442–450.

COHEN, M. R., & NEWSOME, W. T. (2004). What electrical microstimulation has revealed about the neural basis of cognition. *Curr. Opin. Neurobiol.*, *14*, 169–177.

CUMMING, B. G., & DEANGELIS, G. C. (2001). The physiology of stereopsis. *Annu. Rev. Neurosci.*, *24*, 203–238.

CUMMING, B. G., & PARKER, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, *389*, 280–283.

CUMMING, B. G., & PARKER, A. J. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J. Neurosci.*, *19*, 5602–5618.

CUMMING, B. G., & PARKER, A. J. (2000). Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *J. Neurosci.*, *20*, 4758–4767.

DEANGELIS, G. C. (2000). Seeing in three dimensions: The neurophysiology of stereopsis. *Trends Cogn. Sci.*, *4*, 80–90.

DEANGELIS, G. C., CUMMING, B. G., & NEWSOME, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature*, *394*, 677–680.

DEANGELIS, G. C., & NEWSOME, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *J. Neurosci.*, *19*, 1398–1415.

DEANGELIS, G. C., & UKA, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J. Neurophysiol.*, *89*, 1094–1111.

DURAND, J. B., NELISSEN, K., JOLY, O., WARDAK, C., TODD, J. T., NORMAN, J. F., JANSSEN, P., VANDUFFEL, W., & ORBAN, G. A. (2007). Anterior regions of monkey parietal cortex process visual 3D shape. *Neuron*, *55*, 493–505.

EIFUKU, S., & WURTZ, R. H. (1999). Response to motion in extrastriate area MSTl: Disparity sensitivity. *J. Neurophysiol.*, *82*, 2462–2475.

FERRAINA, S., PARE, M., & WURTZ, R. H. (2000). Disparity sensitivity of frontal eye field neurons. *J. Neurophysiol.*, *83*, 625–629.

FREEMAN, R. D. (2004). Binocular interaction in the visual cortex. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 765–778). Cambridge, MA: MIT Press.

FREEMAN, R. D., & OHZAWA, I. (1990). On the neurophysiological organization of binocular vision. *Vis. Res.*, *30*, 1661–1676.

GENOVESIO, A., & FERRAINA, S. (2004). Integration of retinal disparity and fixation-distance related signals toward an egocentric coding of distance in the posterior parietal cortex of primates. *J. Neurophysiol.*, *91*, 2670–2684.

GNADT, J. W., & MAYS, L. E. (1995). Neurons in monkey parietal area LIP are tuned for eye-movement parameters in three-dimensional space. *J. Neurophysiol.*, *73*, 280–297.

GONZALEZ, F., & PEREZ, R. (1998). Neural mechanisms underlying stereoscopic vision. *Prog. Neurobiol.*, *55*, 191–224.

GRAHAM, M. E., & ROGERS, B. J. (1982). Simultaneous and successive contrast effects in the perception of depth from motion-parallax and stereoscopic information. *Perception*, *11*, 247–262.

HEGDE, J., & VAN ESSEN, D. C. (2005). Role of primate visual area V4 in the processing of 3-D shape characteristics defined by disparity. *J. Neurophysiol.*, *94*, 2856–2866.

HINKLE, D. A., & CONNOR, C. E. (2002). Three-dimensional orientation tuning in macaque area V4. *Nat. Neurosci.*, *5*, 665–670.

HINKLE, D. A., & CONNOR, C. E. (2005). Quantitative characterization of disparity tuning in ventral pathway area V4. *J. Neurophysiol.*, *94*, 2726–2737.

HOWARD, I. P., & ROGERS, B. J. (1995). *Binocular vision and stereopsis*. New York: Oxford University Press.

HOWARD, I. P., & ROGERS, B. J. (2002). *Seeing in depth. Vol. 2: Depth perception*. Toronto: I. Porteus.

JANSSEN, P., VOGELS, R., LIU, Y., & ORBAN, G. A. (2001). Macaque inferior temporal neurons are selective for three-dimensional boundaries and surfaces. *J. Neurosci.*, *21*, 9419–9429.

JANSSEN, P., VOGELS, R., LIU, Y., & ORBAN, G. A. (2003). At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron*, *37*, 693–701.

JANSSEN, P., VOGELS, R., & ORBAN, G. A. (1999). Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proc. Natl. Acad. Sci. USA*, *96*, 8217–8222.

JANSSEN, P., VOGELS, R., & ORBAN, G. A. (2000a). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science*, *288*, 2054–2056.

JANSSEN, P., VOGELS, R., & ORBAN, G. A. (2000b). Three-dimensional shape coding in inferior temporal cortex. *Neuron*, *27*, 385–397.

JULESZ, B. (1971/1997). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.

KRUG, K. (2004). A common neuronal code for perceptual processes in visual cortex? Comparing choice and attentional correlates in V5/MT. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *359*, 929–941.

LI, C. Y., LEI, J. J., & YAO, H. S. (1999). Shift in speed selectivity of visual cortical neurons: A neural basis of perceived motion contrast. *Proc. Natl. Acad. Sci. USA*, *96*, 4052–4056.

MAUNSELL, J. H., & VAN ESSEN, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque

monkey: I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.*, *49*, 1127–1147.

Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey: II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.*, *49*, 1148–1167.

Nadler, J. W., Angelaki, D. E., & DeAngelis, G. C. (2006). A smooth eye movement signal provides the extraretinal input used by MT neurons to code depth sign from motion parallax. *Soc. Neurosci. Abstr.*, *407*, 8.

Nadler, J. W., Angelaki, D. E., & DeAngelis, G. C. (2008). A neural representation of depth from motion parallax in macaque visual cortex. *Nature*, *452*, 642–645.

Naji, J. J., & Freeman, T. C. (2004). Perceiving depth order during pursuit eye movement. *Vis. Res.*, *44*, 3025–3034.

Nawrot, M. (2003a). Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. *Vis. Res.*, *43*, 1553–1562.

Nawrot, M. (2003b). Depth from motion parallax scales with eye movement gain. *J. Vis.*, *3*, 841–851.

Nawrot, M., & Joyce, L. (2006). The pursuit theory of motion parallax. *Vis. Res.*, *46*, 4709–4725.

Neri, P. (2005). A stereoscopic look at visual cortex. *J. Neurophysiol.*, *93*, 1823–1826.

Neri, P., Bridge, H., & Heeger, D. J. (2004). Stereoscopic processing of absolute and relative disparity in human visual cortex. *J. Neurophysiol.*, *92*, 1880–1891.

Nguyenkim, J. D., & DeAngelis, G. C. (2003). Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J. Neurosci.*, *23*, 7117–7128.

Nguyenkim, J. D., & DeAngelis, G. C. (2004). Macaque MT neurons are selective for 3D surface orientation defined by multiple cues. *Soc. Neurosci. Abstr.*, *368*, 12.

Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2004). Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *J. Neurosci.*, *24*, 2065–2076.

Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2005). Neuronal computation of disparity in V1 limits temporal resolution for detecting disparity modulation. *J. Neurosci.*, *25*, 10207–10219.

Nienborg, H., & Cumming, B. G. (2006). Macaque V2 neurons, but not V1 neurons, show choice-related activity. *J. Neurosci.*, *26*, 9567–9578.

Nienborg, H., & Cumming, B. G. (2007). Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. *Nat. Neurosci.*, *10*, 1608–1614.

Nover, H., Anderson, C. H., & DeAngelis, G. C. (2005). A logarithmic, scale-invariant representation of speed in macaque area MT accounts for speed discrimination performance. *J. Neurosci.*, *25*, 10049–10060.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1997). The neural coding of stereoscopic depth. *NeuroReport*, *8*, iii–xii.

Palanca, B. J., & DeAngelis, G. C. (2003). Macaque middle temporal neurons signal depth in the absence of motion. *J. Neurosci.*, *23*, 7647–7658.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nat. Rev. Neurosci.*, *8*, 379–391.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annu. Rev. Neurosci.*, *21*, 227–277.

Poggio, G. F., & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J. Neurophysiol.*, *40*, 1392–1405.

Ponce, C. R., Lomber, S. G., & Born, R. T. (2008). Integrating motion and depth via parallel pathways. *Nat. Neurosci.*, *11*, 216–223.

Prince, S. J., Pointon, A. D., Cumming, B. G., & Parker, A. J. (2000). The precision of single neuron responses in cortical area V1 during stereoscopic depth judgments. *J. Neurosci.*, *20*, 3387–3400.

Prince, S. J., Pointon, A. D., Cumming, B. G., & Parker, A. J. (2002). Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. *J. Neurophysiol.*, *87*, 191–208.

Rodman, H. R., & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the macaque. *Vis. Res.*, *27*, 2035–2048.

Rogers, B. J. (1993). Motion parallax and other dynamic cues for depth in humans. *Rev. Oculomot. Res.*, *5*, 119–137.

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*, 125–134.

Rogers, B. J., & Graham, M. E. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vis. Res.*, *22*, 261–270.

Rogers, S., & Rogers, B. J. (1992). Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Percept. Psychophys.*, *52*, 446–452.

Roy, J. P., Komatsu, H., & Wurtz, R. H. (1992). Disparity sensitivity of neurons in monkey extrastriate area MST. *J. Neurosci.*, *12*, 2478–2492.

Salzman, C. D., Murasugi, C. M., Britten, K. H., & Newsome, W. T. (1992). Microstimulation in visual area MT: Effects on direction discrimination performance. *J. Neurosci.*, *12*, 2331–2355.

Sedgwick, H. A. (1986). Space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 1. Sensory processes and perception* (pp. 1–57). New York: John Wiley.

Taira, M., Tsutsui, K. I., Jiang, M., Yara, K., & Sakata, H. (2000). Parietal neurons represent surface orientation from the gradient of binocular disparity. *J. Neurophysiol.*, *83*, 3140–3146.

Takemura, A., Inoue, Y., Kawano, K., Quaia, C., & Miles, F. A. (2001). Single-unit activity in cortical area MST associated with disparity-vergence eye movements: Evidence for population coding. *J. Neurophysiol.*, *85*, 2245–2266.

Tanabe, S., Doi, T., Umeda, K., & Fujita, I. (2005). Disparity-tuning characteristics of neuronal responses to dynamic random-dot stereograms in macaque visual area V4. *J. Neurophysiol.*, *94*, 2683–2699.

Tehovnik, E. J. (1996). Electrical stimulation of neural tissue to evoke behavioral responses. *J. Neurosci. Methods*, *65*, 1–17.

Thomas, O. M., Cumming, B. G., & Parker, A. J. (2002). A specialization for relative disparity in V2. *Nat. Neurosci.*, *5*, 472–478.

Treue, S., & Andersen, R. A. (1996). Neural responses to velocity gradients in macaque cortical area MT. *Visual Neurosci.*, *13*, 797–804.

Tsutsui, K., Jiang, M., Yara, K., Sakata, H., & Taira, M. (2001). Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. *J. Neurophysiol.*, *86*, 2856–2867.

Tsutsui, K., Sakata, H., Naganuma, T., & Tanaka, Y. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science*, *298*, 409–412.

Uka, T., & DeAngelis, G. C. (2003). Contribution of middle temporal area to coarse depth discrimination: Comparison of

neuronal and psychophysical sensitivity. *J. Neurosci.*, *23*, 3515–3530.

Uka, T., & DeAngelis, G. C. (2004). Contribution of area MT to stereoscopic depth perception: Choice-related response modulations reflect task strategy. *Neuron*, *42*, 297–310.

Uka, T., & DeAngelis, G. C. (2006). Linking neural representation to function in stereoscopic depth perception: Roles of the middle temporal area in coarse versus fine disparity discrimination. *J. Neurosci.*, *26*, 6791–6802.

Uka, T., Tanaka, H., Yoshiyama, K., Kato, M., & Fujita, I. (2000). Disparity selectivity of neurons in monkey inferior temporal cortex. *J. Neurophysiol.*, *84*, 120–132.

Umeda, K., Tanabe, S., & Fujita, I. (2007). Representation of stereoscopic depth based on relative disparity in macaque area V4. *J. Neurophysiol.*, *98*, 241–252.

Westheimer, G. (1979). Cooperative neural processes involved in stereoscopic acuity. *Exp. Brain Res.*, *36*, 585–597.

Xiao, D. K., Marcar, V. L., Raiguel, S. E., & Orban, G. A. (1997). Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. *Eur. J. Neurosci.*, *9*, 956–964.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *J. Physiol.*, *236*, 549–573.

34 Multisensory Integration for Heading Perception in Macaque Visual Cortex

DORA E. ANGELAKI, YONG GU, AND GREGORY C. DEANGELIS

ABSTRACT The brain combines different sources of sensory information to optimize perception. Information from different sensory modalities is often seamlessly integrated into a unified percept with improved behavioral performance. Here we summarize the first attempt to understand the neural basis of multisensory cue integration in the context of a behavioral task in which cues are combined according to statistically optimal predictions. We describe multisensory cue integration in the macaque extrastriate visual cortex using a simple heading discrimination task in which monkeys were asked to judge their direction of self-motion using visual (optic flow) and extraretinal (vestibular) cues. Results suggest that rhesus macaques and humans use similar computational principles for combining multiple sensory cues and that these principles can be accounted for by the properties of individual neurons in multisensory cortical areas.

A fundamental aspect of our sensory experience is that information from different modalities is often seamlessly integrated into a unified percept. Examples of multisensory cue integration include a number of well-known sensory illusions, such as the McGurk effect (McGurk & MacDonald, 1976), ventriloquism (Bertelson & Radeau, 1981), and the illusion of self-motion triggered by visual motion, known as *vection* (Previc, 1992). Combining sensory inputs can improve behavioral performance on a number of tasks, including object recognition (Molholm, Ritter, Javitt, & Foxe, 2004), stimulus detection (Frassinetti, Bolognini, & Ladavas, 2002), and localization (Hairston et al., 2003).

Recently, understanding of multisensory integration has gained momentum, as several psychophysical studies have shown that human observers combine sensory cues according to a statistically optimal weighting scheme derived from Bayesian probability theory (Mamassian, Landy, & Moloney, 2002; Kersten, Mamassian, & Yuille, 2004; Knill & Pouget, 2004). The basic concept is that there exists inherent uncertainty in the information that is available to our senses, as well as in the encoding of that information by our sensory apparatus. Consequently, perceptual judgments should rely on computations involving conditional probability density functions, sensory likelihoods, and prior probability functions that are consistent with the Bayesian framework (Clark & Yuille, 1990; Knill & Pouget, 2004). Assuming Gaussian distributions of the underlying sensory information, independent noise sources, and broad prior distributions relative to the individual cue likelihoods, it is predicted that an optimal estimator (in terms of minimizing the variance of the final estimate) will combine sensory information using a rule that weights the cues according to their reliability (Ernst & Banks, 2002; Knill & Saunders, 2003). As a result, weaker cues would have a lower weighting in the bimodal estimate. In addition, the variance of the bimodal estimate (σ^2_{bi}, as assessed by psychophysical performance) should be lower than that of the unimodal estimates, $\sigma^2_{1,2}$, as given by (see figure 34.1)

$$\sigma^2_{bi} = \sigma^2_1 * \sigma^2_2/(\sigma^2_1 + \sigma^2_2) \quad (1)$$

These predictions have been tested in human psychophysical experiments using a number of different paradigms (van Beers, Sittig, & Gon, 1999; Ernst & Banks, 2002; Knill & Saunders, 2003; Alais & Burr, 2004; Hillis, Watt, Landy, & Banks, 2004). The basic result is remarkably consistent across studies: When combining multiple sensory cues, humans perform as nearly optimal Bayesian observers. Yet no direct neural correlates of these phenomena have been available, in part owing to the lack of a suitable animal model for combined behavioral and electrophysiological experiments in the context of cue integration. Rather, studies of multisensory integration at the neuronal level have often been performed in either anesthetized or passively fixating animals, and the pioneering studies of Stein and colleagues emphasized nonlinearity (superadditivity) as the hallmark of multisensory integration (Meredith & Stein, 1983, 1986, 1996; Wallace, Wilkinson, & Stein, 1996).

DORA E. ANGELAKI and YONG GU Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri

GREGORY C. DEANGELIS Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, New York

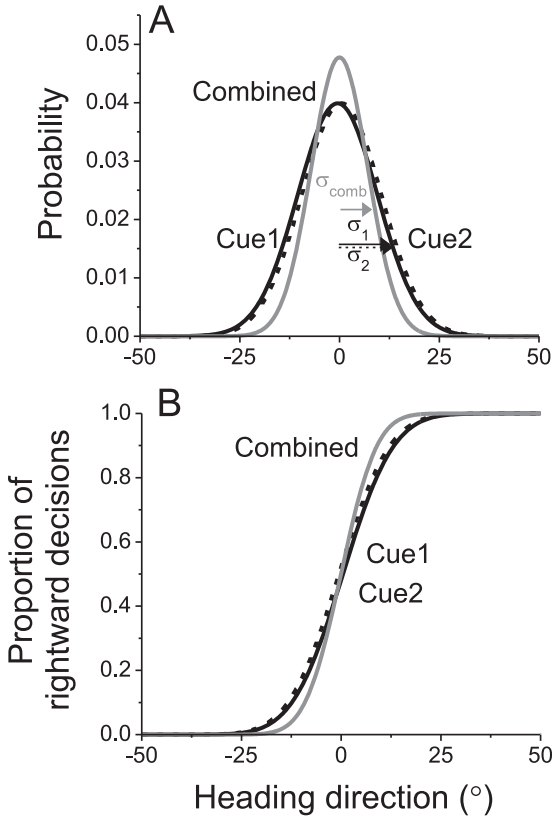


FIGURE 34.1 Schematic illustration of one of the predictions of optimal cue integration. (*A*) Probability density functions (sensory likelihoods) corresponding to two cues: cue 1 (solid curve, e.g., vestibular) and cue 2 (dashed curve, e.g., visual). It is predicted that the bimodal probability distribution (gray curve) will be narrower than those for the individual cues (equation 1). This improvement will be largest when the two single cues have the same standard deviation (σ). (*B*) Expected performance for an ideal observer judging heading on the basis of the probability distributions in panel *A*. In this case, the threshold in the combined (bimodal) condition is predicted to be lower than both single-cue thresholds. (Modified with permission from Ernst & Banks, 2002.)

Here we summarize the first attempt to understand the neural basis of multisensory cue integration in the context of a behavioral task in which cues are combined according to the statistically optimal predictions. We describe multisensory cue integration in the macaque extrastriate visual cortex using a simple heading discrimination task in which monkeys are asked to judge their direction of self-motion using visual (optic flow) and extraretinal (vestibular) cues (figure 34.2*A*).

Perception of heading from optic flow and vestibular signals

How do we perceive our direction of self-motion through space? To navigate effectively through a complex three-dimensional environment, we must accurately estimate our own motion relative to objects around us. Self-motion perception is an intriguing problem in sensory integration, requiring the neural combination of visual signals (e.g., optic flow), vestibular signals regarding head motion, and perhaps somatosensory and proprioceptive cues (Dichgans & Brandt, 1974; Hlavacka, Mergner, & Schweigart, 1992; Hlavacka, Mergner, & Bolha, 1996). In particular, patterns of image motion across the retina (optic flow) can provide strong cues to self-motion, as is evidenced by the fact that optic flow alone can elicit the illusion of self-motion. As early as 1875, Ernst Mach described self-motion sensations (i.e., circular and linear vection) induced by visual stimuli. Since then, several other studies have characterized the behavioral observation that large-field optic flow stimulation induces self-motion perception (Brandt, Dichgans, & Koenig, 1973; Berthoz, Pavard, & Young, 1975; Dichgans & Brandt, 1978). Although self-motion perception generally involves the analysis of observer translation and rotation, we shall limit our scope here to translational movements. Thus the central issue we explore is how we compute our direction of heading.

Many visual psychophysical and theoretical studies have shown that optic flow provides powerful cues to heading (Gibson, 1950) and have examined how heading can be computed from optic flow (Warren, 2003). In parallel, independent information about the motion of our head or body in space can arise from the vestibular system. Specifically, vestibular signals provide information about the angular and linear accelerations of the head in space (Angelaki, 2004; Angelaki & Cullen, 2008) and thus provide important cues to self-motion estimation. A role of the vestibular system in the perception of self-motion has long been acknowledged (Guedry, 1974, 1978; Benson, Spencer, & Stott, 1986; Telford, Howard, & Ohmi, 1995).

In one such heading discrimination task, the subject experiences forward motion with a small leftward or rightward component. At the end of each trial, the task requires an eye movement to report whether the subject experienced leftward or rightward motion (figure 34.2*A*). Both humans (Smith, Bush, & Stone, 2002) and monkeys (Gu, DeAngelis, & Angelaki, 2007) can be quite accurate in discriminating their heading direction in the absence of optic flow, with thresholds that can be as small as 1–3 degrees during motion in darkness. These threshold values during motion in darkness are comparable to (although larger than) those described in visual heading discrimination tasks (Warren & Hannon, 1990; Royden, Banks, & Crowell, 1992; van den Berg & Brenner, 1994; Stone & Perrone, 1997). Vestibular heading thresholds increase more than ten-fold after bilateral labyrinthectomy (figure 34.2*B*, solid symbols), suggesting that vestibular information is critical for heading discrimination. Although some recovery was seen over the first few days postlesion, thresholds remained elevated when measured 3–6 months following the lesion (Gu et al., 2007). In contrast, labyrinthectomy had a very modest effect on visual

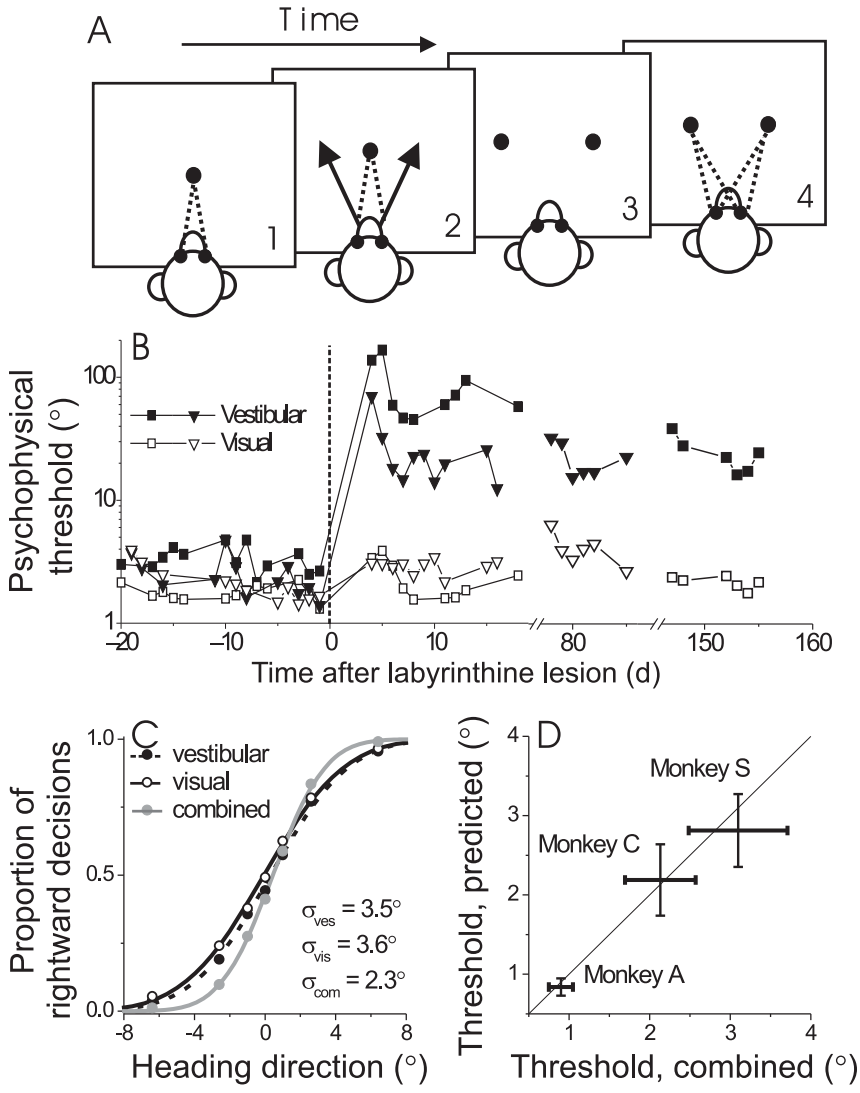


FIGURE 34.2 Heading discrimination task and behavioral performance. (*A*) Task layout. Monkeys were seated on a motion platform and were translated within the horizontal plane to provide vestibular stimulation. A projector mounted on the platform displayed images of a three-dimensional star field and thus provided optic flow information. After fixating a visual target, the monkey experienced forward motion with a small leftward or rightward component and subsequently reported his perceived heading ("left" versus "right") by making a saccadic eye movement to one of two targets. (*B*) Daily psychophysical thresholds before and after bilateral labyrinthectomy (0 marks the day of surgery). Data obtained during the heading discrimination task in the absence of optic flow (solid symbols) are compared with those obtained when the heading was defined exclusively by optic flow (open symbols). Data from two animals are shown with different symbols (Gu et al., 2007). (*C*) Comparison of behavioral performance in one animal under unimodal (vestibular: dashed curve, visual: solid curve) and bimodal (gray curve) conditions. Notice the steeper slope of the psychometric function in the combined condition as compared to the single-cue (visual and vestibular) conditions. (*D*) Comparison of the average psychophysical threshold obtained in the combined condition with that predicted based on statistically optimal cue integration. Data are shown for three animals. (Parts *A* and *B* replotted from Gu et al., 2007.)

heading thresholds where the animals remained stationary and heading was specified solely by optic flow (figure 34.2*B*, open symbols).

To test whether macaques, like humans, combine sensory cues according to a statistically optimal, Bayesian-style weighting scheme, an experiment was performed in which heading was specified not only by optic flow alone or inertial motion alone, but also by bimodal stimulation when congruent visual and vestibular cues were presented together. Three stimulus conditions were randomly interleaved within a single block of trials: (1) a vestibular condition, in which heading was defined solely by inertial motion cues by translating the animal on a motion platform; (2) a visual condition, in which heading was defined solely by optic flow provided by a projector that was mounted on the platform; and (3) a combined condition consisting of congruent inertial

motion and optic flow cues. Each movement trajectory, either real (vestibular condition) or visually simulated (visual condition), had a duration of 2 s and consisted of a Gaussian velocity profile (for details, see Gu, Watkins, Angelaki, & DeAngelis, 2006).

Average behavior, in the form of psychometric functions, from one of the animals is illustrated in figure 34.2*C*. Note that the reliability of the individual cues was roughly equated during training by reducing the coherence of the visual motion stimulus such that visual and vestibular thresholds were approximately equal (figure 34.2*C*, open/solid circles and solid/dashed curves). This balancing of the two cues is crucial, as it affords the maximal opportunity to observe improvement in performance under cue combination (Ernst & Banks, 2002). In the combined condition (figure 34.2*C*, gray circles and curve), the monkey's heading threshold was substantially smaller, as evidenced by the steeper slope of the gray curve. If the monkey combined the two cues optimally, as predicted by Bayesian cue integration principles, thresholds should be reduced by approximately 30% under cue combination (equation 1). That bimodal behavioral thresholds are similar to the optimal cue integration predictions is illustrated in figure 34.2*D* for data from three animals. Thus, like humans, macaques can combine multiple sensory cues nearly optimally to improve perceptual performance.

This demonstration of near-optimal cue integration in the monkey's behavior provides a unique opportunity to search for the neural basis of Bayesian inference at the level of individual neurons and populations of neurons. In identifying candidate populations of neurons that integrate visual and vestibular signals for self-motion perception, we seek neurons that are tuned for direction of motion in optic flow fields and that carry vestibular signals related to the direction of head motion through space. As will be summarized next, such visual/vestibular convergence occurs in multiple cortical areas. In contrast, responsiveness to optic flow is generally absent in subcortical areas with vestibular-related activities, including the brain stem vestibular and deep cerebellar nuclei (Bryan, Meng, DeAngelis, & Angelaki, 2007) and primate thalamus (Meng, May, Dickman, & Angelaki, 2007). In the following, we first briefly summarize what has been previously known regarding visual/vestibular convergence in the macaque cortex; we then describe in more detail how MSTd neurons respond in the context of the multimodal heading discrimination task.

Responses of primate cortical neurons to optic flow and vestibular stimuli

Optic flow-sensitive neurons have been found in the dorsal portion of the medial superior temporal area (MSTd) (Tanaka et al., 1986; Duffy & Wurtz, 1991, 1995), ventral intraparietal area (VIP) (Schaafsma & Duysens, 1996; Bremmer, Duhamel, Ben Hamed, & Graf, 2002; Bremmer, Klam, Duhamel, Ben Hamed, & Graf, 2002), posterior parietal cortex (7a) (Siegel & Read, 1997), and superior temporal polysensory area (STP) (Anderson & Siegel, 1999). In particular, neurons in MSTd/VIP have large visual receptive fields and are selective for optic flow patterns similar to those seen during self-motion (MSTd: Tanaka et al., 1986; Tanaka, Fukada, & Saito, 1989; Duffy & Wurtz, 1991, 1995; Bradley, Maxwell, Anderson, Banks, & Shenoy, 1996; Lappe, Bremmer, Pekel, Thiele, & Hoffmann, 1996); (VIP: Schaafsma & Duysens, 1996; Bremmer, Duhamel, et al., 2002). Importantly, electrical stimulation of MSTd or VIP has been reported to bias heading judgments that are based solely on optic flow (Britten & van Wezel, 1998, 2002; Zhang & Britten, 2003). MSTd/VIP neurons are also selective for motion in darkness, suggesting that they receive vestibular inputs (Duffy, 1998; Bremmer, Kubischek, Pekel, Lappe, & Hoffmann, 1999; Bremmer, Duhamel, et al., 2002; Schlack, Hoffmann, & Bremmer, 2002; Gu et al., 2006; Chen, Henry, DeAngelis, & Angelaki, 2007; Takahashi et al., 2007).

Using a custom-built virtual reality system, the heading selectivity of MSTd (Gu et al., 2006; Takahashi et al., 2007) and VIP (Chen et al., 2007) neurons has recently been quantified in three dimensions. Inertial motion (vestibular) signals were provided by translating a motion platform, and optic flow (visual) signals were provided by a projector that was mounted on the platform and rear-projected images onto a screen in front of the monkey. Approximately 60% of MSTd neurons were significantly tuned for heading under both the visual and vestibular stimulus conditions. These convergent MSTd cells fell into one of two groups: (1) "congruent" neurons, which had similar visual/vestibular preferred directions, thus signaled the same motion direction in three-dimensional space under both unimodal stimulus conditions, and (2) "opposite" neurons, which preferred nearly opposite directions of heading under visual and vestibular stimulus conditions (Gu et al., 2006). The response modulation of MSTd neurons during inertial motion (vestibular condition) was indeed shown to be of labyrinthine origin, as MSTd cells were no longer tuned during inertial motion following bilateral labyrithectomy (figure 34.3) (Gu et al., 2007; Takahashi et al., 2007).

Notably, responsiveness to both visual (optic flow) and vestibular stimulation is generally not present within more traditionally considered areas of "vestibular cortex" (Fredrickson & Rubin, 1986; Fukushima, 1997; Guldin & Grusser, 1998). Three main cortical areas have been characterized as either exhibiting responses to vestibular stimulation and/or receiving short-latency vestibular signals (trisynaptic through the vestibular nuclei and the thalamus). They are (1) area 2v, located in the transition zone of areas 2, 5, and 7 within the intraparietal sulcus (Fredrickson,

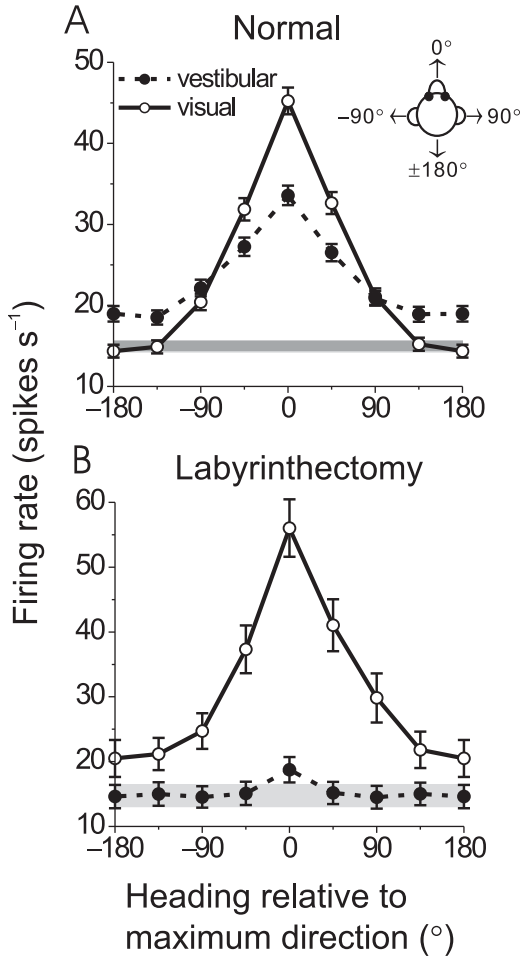


FIGURE 34.3 Comparison of visual and vestibular population tuning curves (*A*) before and (*B*) after labyrinthectomy. Visual responses (open circles and solid lines) of each neuron were shifted to align the peaks of all tuning curves at 0° prior to averaging across the population. Vestibular responses (solid circles and dashed line) were also averaged across neurons after being aligned to the vestibular maximum response direction. Gray bands in both panels indicate the average spontaneous firing rate ± standard error. (Modified from Gu et al., 2007.)

Scheid, Figge, & Kornhuber, 1966; Schwarz & Fredrickson, 1971a, 1971b; Buttner & Buettner, 1978); (2) the parietoinsular vestibular cortex, located between the auditory and secondary somatosensory cortices (Grusser, Pause, & Schreiter, 1990a, 1990b); and (3) area 3a, located within the central sulcus extending into the anterior bank of the precentral gyrus (Odkvist, Schwarz, Frederickson, & Hassler, 1974; Guldin, Akbarian, & Grusser, 1992). Preliminary results suggest that these areas are unlikely to play important roles in visual/vestibular cue integration for heading perception because of the absence of responses to optic flow (Chen et al., 2006; Chen, Henry, DeAngelis, & Angelaki, 2007). Instead, the extrastriate visual cortical areas MSTd and VIP appear to be more likely candidates for visual/vestibular cue integration for heading perception.

Thus, in summary, a handful of studies have examined how cortical neurons integrate visual and vestibular signals to code heading direction. Although these studies clearly establish the presence of both visual and vestibular signals in areas MSTd and VIP, they are limited by the fact that neuronal responses during unimodal or bimodal visual/vestibular stimulation were obtained while monkeys either were passively fixating a target or were freely allowed to make eye movements in darkness. Yet to probe the neural basis of multisensory cue integration and the neural correlates of Bayesian inference, neural activity needs to be measured in the context of a behavioral task that requires the subject to report his/her perception of self-motion. To make concrete links between neural activity and self-motion perception, we must record and/or manipulate neural activity during such tasks. Next we describe such an experiment.

Responses of MSTd neurons during unimodal and bimodal variants of the heading discrimination task

Having established robust cue integration behavior in macaques (figures 34.2*C* and 34.2*D*), we recorded from single neurons in area MSTd while monkeys performed the heading discrimination task. To identify multimodal neurons, we measured heading-tuning curves in the horizontal plane under both single-cue conditions while animals maintained visual fixation. Figures 34.4*A* and 34.4*B* show data from two example multimodal neurons with clear tuning under both single-cue conditions. The neuron in figure 34.4*A* preferred leftward (negative) headings for both stimuli and was classified as a "congruent" cell. In contrast, the neuron in figure 34.4*B* preferred leftward headings under the visual condition (solid line) and rightward headings under the vestibular condition (dashed line) and was classified as an "opposite" cell. (Note that heading directions are referenced to either the real or simulated self-motion; thus similar tuning in the visual and vestibular conditions defines a congruent cell.)

Over the much narrower range of headings sampled during discrimination, the unimodal tuning of these example neurons was monotonic in all three stimulus conditions (figures 34.4*C* and 34.4*D*). For the congruent cell (figure 34.4*C*), heading tuning became steeper in the combined condition. In contrast, the tuning curve became flatter in the combined condition for the opposite cell (figure 34.4*D*). To compare neuronal and behavioral sensitivity more directly, we used signal detection theory (Green & Swets, 1966; Britten, Shadlen, Newsome, & Movshon, 1992) to quantify the ability of an ideal observer to discriminate heading on the basis of the activity of a single neuron (figures 34.4*E* and 34.4*F*, symbols). We fitted these neurometric data with cumulative Gaussian functions (figures 34.4*E* and 34.4*F*, smooth curves) and defined the neuronal threshold as the standard deviation of the Gaussian. The smaller the

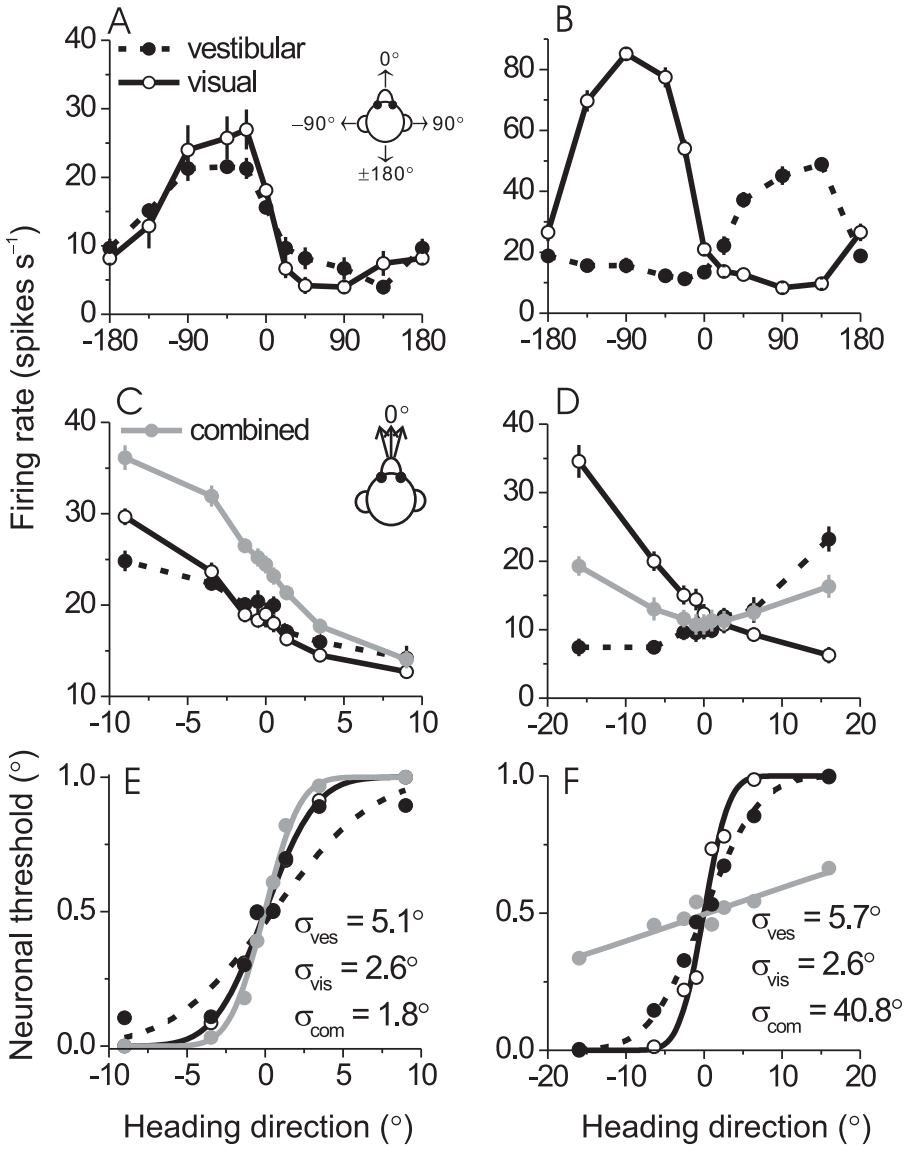


FIGURE 34.4 Heading sensitivity in area MSTd. (*A*, *B*) Heading-tuning curves of two example neurons with (*A*) congruent and (*B*) opposite visual/vestibular heading preferences. Negative angles correspond to leftward headings; positive numbers illustrate rightward directions. (*C*, *D*) Responses of the same neurons to a narrow range of heading stimuli presented while the monkey performed the discrimination task. (*E*, *F*) Neurometric functions computed by ROC analysis for the same two neurons. Smooth curves show best-fitting cumulative Gaussian functions. (Modified from Gu et al., 2008.)

threshold, the steeper is the neurometric function and the more sensitive the neuron is to subtle variations in heading. For the congruent neuron in figure 34.4*E*, the neuronal threshold was smallest in the combined condition (gray symbols and lines), indicating that the neuron could discriminate smaller variations in heading when both cues were provided. In contrast, for the opposite neuron in figure 34.4*F*, the reverse was true: the neuron became less sensitive in the presence of both cues (gray symbols and lines).

The effect of visual/vestibular congruency on neuronal sensitivity during bimodal stimulation held across the whole population of neurons. To summarize this dependency, a quantitative index of visual/vestibular congruency (CI) was established that ranged from +1, when visual and vestibular tuning functions have a consistent slope (figure 34.4*A*), to −1, when they have opposite slopes (figure 34.4*B*). We then computed, for each neuron, the ratio of the neuronal threshold in the combined condition to the threshold expected if neurons combine cues optimally according to equation 1. A significant correlation was seen between the ratio of combined to predicted thresholds and CI (figure 34.5*A*), such that neurons with large positive CIs (congruent cells, black circles) had thresholds close to the optimal

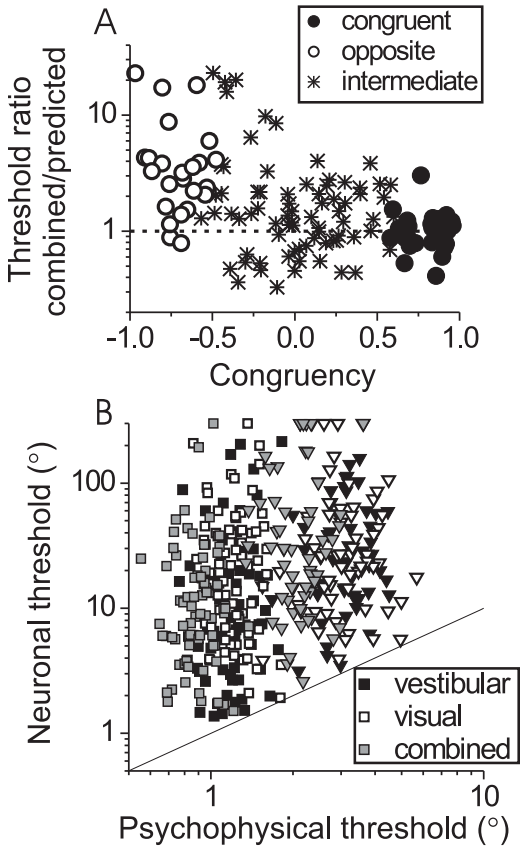


FIGURE 34.5 Neuronal sensitivity of MSTd neurons. (*A*) Neuronal sensitivity under cue combination depends on congruency of visual and vestibular tuning. The ordinate in this scatterplot represents the ratio of the threshold measured in the combined condition to the prediction from optimal cue integration. The abscissa represents the congruency index of heading tuning for visual and vestibular responses (CI). Asterisks denote neurons for which the CI is not significantly different from zero. At the two extremes, neurons with CIs significantly larger than 0 were defined as *congruent* cells (solid symbols), whereas neurons with CIs significantly lower than 0 were defined as *opposite* cells (open symbols). Dashed horizontal line: threshold in the combined condition is equal to the prediction. (*B*) Comparison of neuronal and psychophysical thresholds. Each datum represents one recording session, with solid, open, and gray symbols denoting the vestibular, visual, and combined conditions, respectively (squares and triangles represent data from two animals). Most data points lie well above the diagonal, indicating that most neurons are less sensitive than the monkeys. Only the most sensitive neurons have thresholds comparable to that of the animal. (Modified from Gu et al., 2008.)

prediction (ratios near unity). Thus the average neuronal thresholds for congruent MSTd cells followed a pattern similar to the monkeys' behavior. In contrast, combined thresholds for opposite cells were generally much higher than predicted from optimal cue integration (figure 34.5*A*, open circles), indicating that these neurons became less sensitive during cue combination.

Notably, only the most sensitive neurons rivaled behavioral performance, whereas most neurons were substantially less sensitive than the animal's behavior, and this was true under both unimodal and bimodal stimulation (figure 34.5*B*). To perform the task based on MSTd activity, the monkey must therefore either pool responses across many neurons or rely more heavily on the most sensitive neurons (Parker & Newsome, 1998). Note that the stimulus range in our task was not tailored to the tuning of individual neurons, such that many neurons have large thresholds mainly because the tuning curve was flat over the range of headings tested during discrimination.

Correlations with behavioral choice

If monkeys rely on area MSTd for heading discrimination, the results of figure 34.5*A* suggest that they selectively monitor the activity of congruent cells and not opposite cells. To test this hypothesis, we computed choice probabilities (CPs) (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996) to quantify whether trial-to-trial fluctuations in neural firing rates were correlated with fluctuations in the monkeys' perceptual decisions (for a constant physical stimulus). A significant CP greater than 0.5 indicates that the monkey tends to choose the neuron's preferred sign of heading (leftward versus rightward) when the neuron fires more strongly. Such a result is thought to reflect a functional link between the neuron and perception (Britten et al., 1996; Parker & Newsome, 1998; Krug, 2004). Notably, although MSTd is classically considered visual cortex, vestibular CPs were significantly larger than chance (Gu et al., 2007). Moreover, vestibular signals were consistently correlated with heading percepts irrespective of congruency (figure 34.6*A*).

In contrast, perhaps surprisingly, CPs were overall smaller under the visual and combined conditions (0.52 in both cases, compared to 0.55 for the vestibular condition). These CPs, when averaged across the whole MSTd population, are small because, like neuronal thresholds, CPs in the visual and combined conditions depend on congruency (figures 34.6*B* and 34.6*C*). Specifically, congruent cells tended to have positive CPs (>0.5), and opposite cells tended to have negative CPs (<0.5) or CPs near zero. Paradoxically, neurons with a significant CP less than 0.5 increase their firing rates when the monkey chooses their nonpreferred direction. Note that because the description of visual CPs as larger or smaller than 0.5 is based on visual response tuning, opposite MSTd neurons would consistently have positive (>0.5) CPs when expressed relative to the cell's vestibular preference. This finding suggests that visual responses may be decoded relative to the vestibular preference of the neurons. Perhaps most important, congruent cells were much more strongly correlated with monkeys' heading judgments in the combined condition than were opposite cells, consistent with the idea that the animals might have selectively monitored congruent cells to achieve near-optimal cue integration.

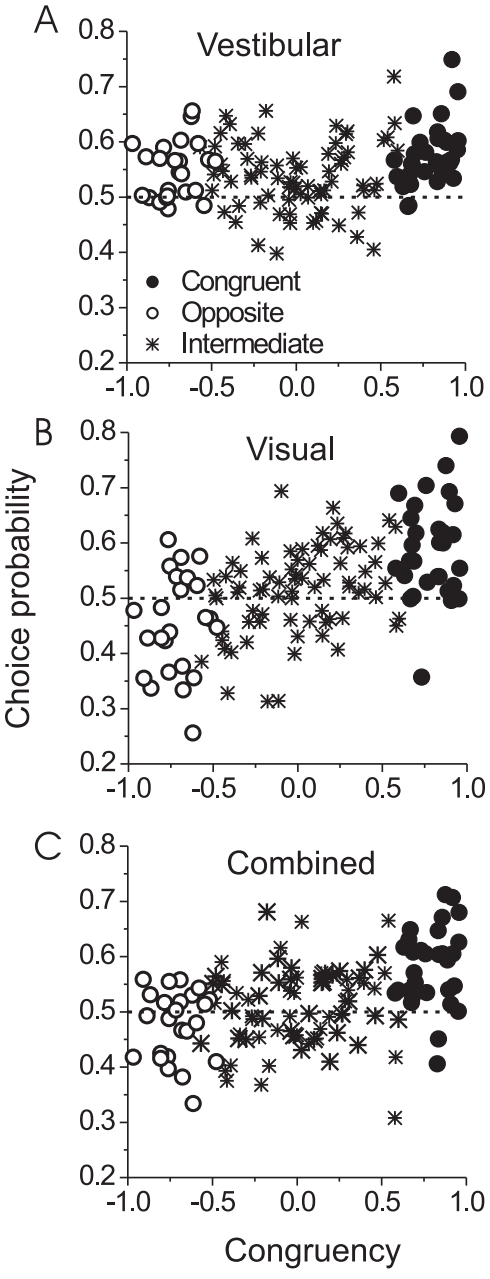


FIGURE 34.6 Correlations between MSTd responses and perceptual decisions depend on congruency of tuning. Choice probability (CP) data are plotted as a function of congruency index (CI) for each MSTd neuron tested in the (*A*) vestibular, (*B*) visual, and (*C*) combined conditions. Congruent, opposite, and intermediate neurons are classified as in figure 34.5. (Modified from Gu et al., 2008.)

In summary, by simultaneously monitoring neural activity and behavior, it has been possible to study neural mechanisms of multisensory processing under conditions in which cue integration is known to take place perceptually. In addition to demonstrating near-optimal cue integration by monkeys, a population of neurons has been identified in area MSTd that could account for improvement in psychophysical performance under cue combination. These findings implicate area MSTd in sensory integration for heading perception and establish an excellent model system for studying the detailed mechanisms by which neurons combine different sensory signals and dynamically reweight these signals to optimize performance as the reliability of cues varies (Knill & Pouget, 2004). However, because the reliability of the visual and vestibular cues was not varied in these experiments, it is currently unclear whether monkeys and MSTd neurons dynamically reweight these cues, as predicted by statistically optimal cue integration schemes. While experiments are currently underway to test this very important prediction of Bayesian cue integration in trained animals (Fetsch, Angelaki, & DeAngelis, 2007), we next summarize results from a simpler experiment in which the reliability of the visual cue was varied during neural recordings in a passively fixating animal (Morgan, DeAngelis, & Angelaki, 2008). This experiment sought to characterize the mathematical rule by which MSTd neurons combine their visual and vestibular inputs. Specifically, we asked whether bimodal responses in MSTd are well fit by weighted linear sums of unimodal responses or whether a nonlinear combination rule is required. Moreover, we asked whether the weights that neurons apply to these cues change with the relative reliabilities of the two cues.

Dependence on cue reliability

As a first step to investigate how cue reliability modulates visual/vestibular cue integration, we compared unimodal responses to eight evenly spaced directions (45 degrees apart) in the horizontal plane with bimodal responses to the 64 possible combinations of these eight vestibular and visual headings, including eight congruent and 56 incongruent (cue-conflict) presentations. In all conditions, monkeys were simply required to maintain fixation during stimulus presentation. For each unimodal stimulus (visual and vestibular), tuning curves were constructed by plotting the mean response versus heading direction, as illustrated for an example congruent cell in figure 34.7 (tuning curves along the left ordinate and abscissa). For the bimodal stimuli, in which each response is associated with both a vestibular heading and a visual heading, responses have been illustrated as two-dimensional contour maps with vestibular heading along the abscissa and visual heading along the ordinate (figure 34.7). At 100% visual coherence, bimodal responses often more strongly reflected the visual unimodal tuning preference, as indicated by the horizontal band of high firing rates (figure 34.7*A*).

Lowering visual coherence, by reducing the proportion of dots carrying the motion signal (Takahashi et al., 2007; Morgan et al., 2008), altered both the unimodal visual responses and the pattern of bimodal responses (figures 34.7*B* and 34.7*C*). In both cases, the visual heading tuning

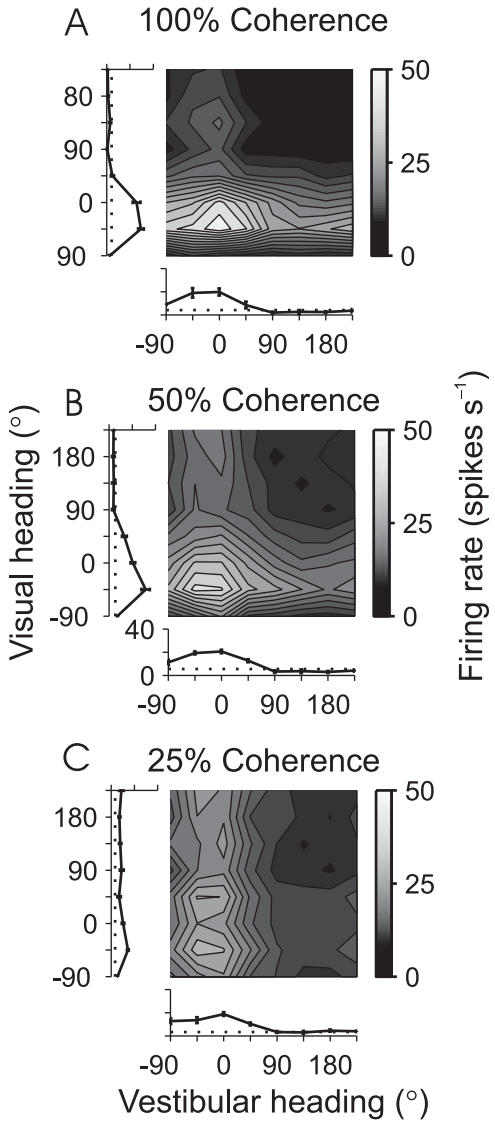


FIGURE 34.7 Comparison of unimodal and bimodal tuning for a congruent MSTd cell, tested at three motion coherences. Grayscale maps show mean firing rates as a function of vestibular and visual headings in the bimodal condition (including all 64 possible combinations of eight visual headings and eight vestibular headings). Tuning curves along the left and bottom margins show mean (± standard error of the mean) firing rates versus heading for the unimodal visual and vestibular conditions, respectively. (*A*) Bimodal responses at 100% coherence are visually dominated. (*B*) Bimodal responses at 50% coherence show a balanced contribution of visual and vestibular cues. (*C*) At 25% coherence, bimodal responses appear to be dominated by the vestibular input. (Replotted with permission from Morgan et al., 2008.)

(tuning curve along ordinate) remained similar in shape and heading preference, but the peak-to-trough response modulation was reduced at 50% and 25% coherence. Simultaneously, the transition from visual dominance to vestibular dominance is clearly evident in the bimodal responses as a function of coherence. Although visually dominated at 100% coherence (horizontal band in figure 34.7*A*), bimodal responses became progressively more influenced by the vestibular cue as coherence was reduced. At 50% coherence, the presence of a clear symmetric peak suggests well-matched visual and vestibular contributions to the bimodal response (figure 34.7*B*). As visual coherence was further reduced to 25%, vestibular dominance is observed, the bimodal response taking the form of a vertical band aligned with the vestibular heading preference (figure 34.7*C*).

Bimodal responses were adequately fit by a weighted linear sum of responses from the vestibular and visual conditions, with weights w_{visual} and $w_{vestibular}$ describing the strength of the contributions of each unimodal input to the bimodal response (Morgan et al., 2008). Given that the relative influences of the two cues on the bimodal response changes with motion coherence, as shown in figure 34.7, an important question then arises: Do the weights with which each neuron combines its vestibular and visual inputs remain fixed as coherence changes, the decreased visual influence in the bimodal tuning being simply due to the weaker visual responses at lower coherences? Alternatively, do the weights given to the vestibular and visual inputs change with the relative reliabilities of the two cues? In the former scenario, the multisensory combination rule used by MSTd neurons is independent of cue reliability, whereas in the latter scenario, neurons modify their combination rules when the quality of the sensory cues changes. Quantitative analyses support the latter possibility. In particular, as compared to 100% coherence, vestibular weights at 50% coherence shifted toward larger values (figure 34.8*A*), and visual weights shifted toward smaller values (figure 34.8*B*). This result is further illustrated for neurons recorded at multiple coherences: Visual weights increased, whereas vestibular weights declined with increasing motion coherence (figures 34.8*C* and 34.8*D*).

These results, showing that MSTd neurons give less weight to their visual inputs when optic flow is degraded by reducing motion coherence, might contribute to the observation in human psychophysical studies that the influence of a cue depends on the relative reliability of that cue compared to others (Ernst & Banks, 2002; Battaglia, Jacobs, & Aslin, 2003; Alais & Burr, 2004). However, we cannot yet speak to the temporal dynamics of this reweighting because we presented different motion coherences in separate blocks of trials. Further experiments are necessary to investigate whether neurons reweight their inputs on a trial-by-trial basis.

Conclusion

The past decade has seen a dramatic increase in our understanding of the computational principles that characterize human multisensory perception. Yet little is currently known about the neural mechanisms that underlie probabilistic

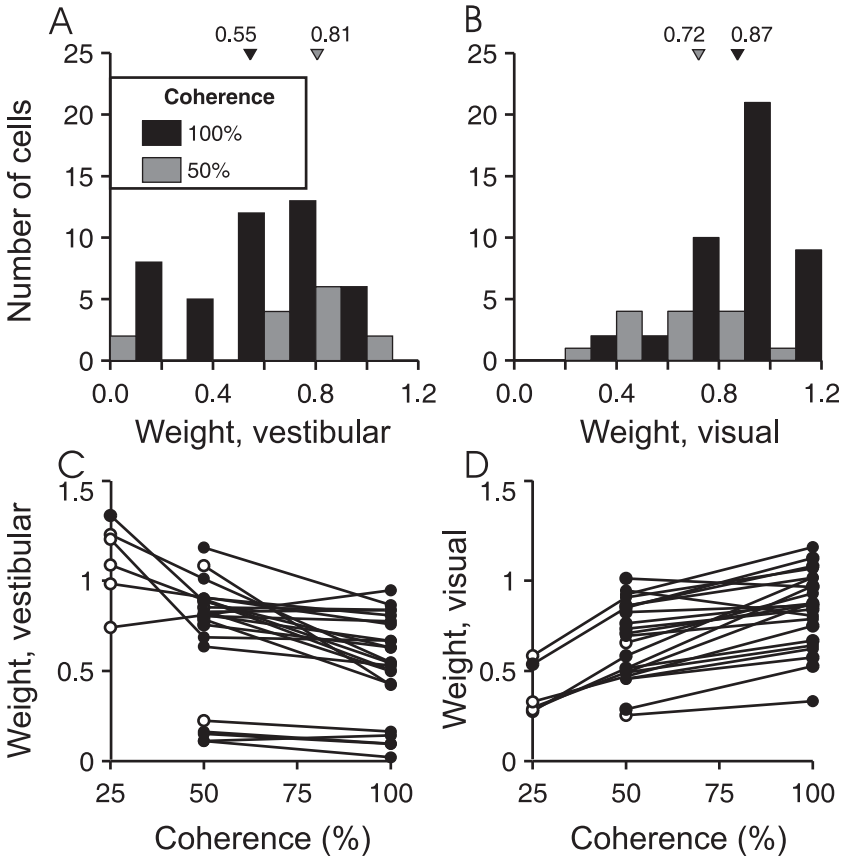


FIGURE 34.8 Dependence of vestibular and visual weights on visual motion coherence. Vestibular and visual weights for each MSTd neuron were derived from linear fits to bimodal responses like those in figure 34.7. (*A*, *B*) Histograms of vestibular and visual weights computed from data at 100% (black) and 50% (gray) coherence. Triangles are plotted at the medians. (*C*, *D*) Vestibular and visual weights are plotted as a function of motion coherence. Data points are coded by the significance of unimodal visual tuning (open versus solid circles). (Replotted with permission from Morgan et al., 2008.)

multisensory integration and Bayesian inference in general. Here, we have summarized recent findings of neurons in extrastriate visual cortex that might mediate visual/vestibular cue integration for heading perception. Although some critical experiments have not yet been conducted, results to date suggest that rhesus macaques and humans use similar computational principles for combining multiple sensory cues and that these principles may be accounted for by the properties of individual neurons in multisensory cortical areas. Multiple questions remain: How distributed are these representations of multisensory integration at the neuronal level? What are the mechanisms by which neurons reweight their inputs according to reliability? Are the responses of sensory cells consistent with the encoding of probability distributions? Finally, how are these sensory signals read out from population responses, and how much of the necessary computations take place in sensory representations versus decision-making networks?

ACKNOWLEDGMENTS This work was supported by NIH EY017866, EY019087, and DC04260 (to D.E.A.) and NIH EY016178 (to G.C.D.). We thank Michael Morgan, whose Ph.D. thesis provided the MSTd data at different visual coherences.

REFERENCES

ALAIS, D., & BURR, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.*, *14*, 257–262.

ANDERSON, K. C., & SIEGEL, R. M. (1999). Optic flow selectivity in the anterior superior temporal polysensory area, STPa, of the behaving monkey. *J. Neurosci.*, *19*, 2681–2692.

ANGELAKI, D. E. (2004). Eyes on target: What neurons must do for the vestibuloocular reflex during linear motion. *J. Neurophysiol.*, *92*, 20–35.

ANGELAKI, D. E., & CULLEN, K. E. (2008). Vestibular system: The many facets of a multimodal sense. *Annu. Rev. Neurosci.*, *31*, 125–150.

BATTAGLIA, P. W., JACOBS, R. A., & ASLIN, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *J. Opt. Soc. Am. [A]*, *20*, 1391–1397.

BENSON, A. J., SPENCER, M. B., & STOTT, J. R. (1986). Thresholds for the detection of the direction of whole-body, linear movement in the horizontal plane. *Aviat. Space Environ. Med.*, *57*, 1088–1096.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophys.*, *29*, 578–584.

Berthoz, A., Pavard, B., & Young, L. R. (1975). Perception of linear horizontal self-motion induced by peripheral vision (linearvection) basic characteristics and visual-vestibular interactions. *Exp. Brain Res.*, *23*, 471–489.

Bradley, D. C., Maxwell, M., Andersen, R. A., Banks, M. S., & Shenoy, K. V. (1996). Mechanisms of heading perception in primate visual cortex. *Science*, *273*, 1544–1547.

Brandt, T., Dichgans, J., & Koenig, E. (1973). Differential effects of central verses peripheral vision on egocentric and exocentric motion perception. *Exp. Brain Res.*, *16*, 476–491.

Bremmer, F., Duhamel, J. R., Ben Hamed, S., & Graf, W. (2002). Heading encoding in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.*, *16*, 1554–1568.

Bremmer, F., Klam, F., Duhamel, J. R., Ben Hamed, S., & Graf, W. (2002). Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.*, *16*, 1569–1586.

Bremmer, F., Kubischik, M., Pekel, M., Lappe, M., & Hoffmann, K. P. (1999). Linear vestibular self-motion signals in monkey medial superior temporal area. *Ann. NY Acad. Sci.*, *871*, 272–281.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.*, *13*, 87–100.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J. Neurosci.*, *12*, 4745–4765.

Britten, K. H., & van Wezel, R. J. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nat. Neurosci.*, *1*, 59–63.

Britten, K. H., & van Wezel, R. J. (2002). Area MST and heading perception in macaque monkeys. *Cereb. Cortex*, *12*, 692–701.

Bryan, A. S., Meng, H., DeAngelis, G. C., & Angelaki, D. E. (2007). Responses of vestibular nucleus cells during visual and vestibular stimulation in three dimensions. *Soc. Neurosci. [Abstracts]*, *180*, 9.

Buttner, U., & Buettner, U. W. (1978). Parietal cortex (2v) neuronal activity in the alert monkey during natural vestibular and optokinetic stimulation. *Brain Res.*, *153*, 392–397.

Chen, A., DeAngelis, G. C., & Angelaki, D. E. (2006). Responses to three-dimensional rotation and translation in the parieto-insular vestibular cortex (PIVC) of rhesus monkey. *Soc. Neurosci. [Abstracts]*, *244*, 6.

Chen, A., Henry, E., DeAngelis, G. C., & Angelaki, D. E. (2007). Comparison of responses to three-dimensional rotation and translation in the ventral intraparietal (VIP) and medial superior temporal (MST) areas of rhesus monkey. *Soc. Neurosci. [Abstracts]*, *715*, 19.

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer Academic.

Dichgans, J., & Brandt, T. (1974). The psychophysics of visually-induced perception of self motion and tilt. In F. O. Schmidt & F. G. Worden (Eds.), *The neurosciences* (pp. 123–129). Cambridge, MA: MIT Press.

Dichgans, J., & Brandt, T. (1978). Visual-vestibular interaction: Effects on self-motion perception and postural control. In R. Held, H. W. Leibowitz, & H. L. Tenber (Eds.), *Handbook of sensory physiology* (Vol. 3, pp. 751–804). New York: Springer.

Duffy, C. J. (1998). MST neurons respond to optic flow and translational movement. *J. Neurophysiol.*, *80*, 1816–1827.

Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli: I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.*, *65*, 1329–1345.

Duffy, C. J., & Wurtz, R. H. (1995). Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *J. Neurosci.*, *15*, 5192–5208.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Fetsch, C. R., Angelaki, D. E., & DeAngelis, G. C. (2007). Dynamic cue re-weighting in rhesus monkeys performing a visual-vestibular heading discrimination task. *Soc. Neurosci. [Abstracts]*, *715*, 26.

Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp. Brain Res.*, *147*, 332–343.

Fredrickson, J. M., & Rubin, A. M. (1986). Vestibular cortex. In E. J. Jones (Ed.), *Cerebral cortex*. (Vol. 5, pp. 99–111). New York: Plenum.

Fredrickson, J. M., Scheid, P., Figge, U., & Kornhuber, H. H. (1966). Vestibular nerve projection to the cerebral cortex of the rhesus monkey. *Exp. Brain Res. 2*, 318–327.

Fukushima, K. (1997). Corticovestibular interactions: Anatomy, electrophysiology, and functional considerations. *Exp. Brain Res.*, *117*, 1–16.

Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley.

Grusser, O. J., Pause, M., & Schreiter, U. (1990a). Localization and responses of neurones in the parieto-insular vestibular cortex of awake monkeys (Macaca fascicularis). *J. Physiol.*, *430*, 537–557.

Grusser, O. J., Pause, M., & Schreiter, U. (1990b). Vestibular neurones in the parieto-insular cortex of monkeys (Macaca fascicularis): Visual and neck receptor responses. *J. Physiol.*, *430*, 559–583.

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multi-sensory cue integration in macaque area MSTd. *Nat. Neurosci.*, *11*(10), 1201–1210.

Gu, Y., DeAngelis, G. C., & Angelaki, D. E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nat. Neurosci.*, *10*, 1038–1047.

Gu, Y., Watkins, P. V., Angelaki, D. E., & DeAngelis, G. C. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J. Neurosci., 26*, 73–85.

Guedry, F. E., Jr. (Ed.). (1974). *Handbook of sensory physiology-vestibular system-psychophysics: Part 2. Applied aspects and general interpretations* (pp. 1–154). Berlin: Springer-Verlag.

Guedry, F. E., Jr. (1978). Visual counteraction on nauseogenic and disorienting effects of some whole-body motions: A proposed mechanism. *Aviat. Space Environ. Med.*, *49*, 36–41.

Guldin, W. O., Akbarian, S., & Grusser, O. J. (1992). Cortico-cortical connections and cytoarchitectonics of the primate vestibular cortex: A study in squirrel monkeys (Saimiri sciureus). *J. Comp. Neurol.*, *326*, 375–401.

Guldin, W. O., & Grusser, O. J. (1998). Is there a vestibular cortex? *Trends Neurosci.*, *21*, 254–259.

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization

ability influences cross-modal bias. *J. Cogn. Neurosci.*, *15*, 20–29.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *J. Vis.*, *4*, 967–992.

Hlavacka, F., Mergner, T., & Bolha, B. (1996). Human self-motion perception during translatory vestibular and proprioceptive stimulation. *Neurosci. Lett.*, *210*, 83–86.

Hlavacka, F., Mergner, T., & Schweigart, G. (1992). Interaction of vestibular and proprioceptive inputs for human self-motion perception. *Neurosci. Lett.*, *138*, 161–164.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol.*, *55*, 271–304.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci.*, *27*, 712–719.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vis. Res.*, *43*, 2539–2558.

Krug, K. (2004). A common neuronal code for perceptual processes in visual cortex? Comparing choice and attentional correlates in V5/MT. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *359*, 929–941.

Lappe, M., Bremmer, F., Pekel, M., Thiele, A., & Hoffmann, K. P. (1996). Optic flow processing in monkey STS: A theoretical and experimental approach. *J. Neurosci.*, *16*, 6265–6285.

Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain* (pp. 13–36). Cambridge, MA: MIT Press.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Meng, H., May, P. J., Dickman, J. D., & Angelaki, D. E. (2007). Vestibular signals in primate thalamus: Properties and origins. *J. Neurosci.*, *27*, 13590–13602.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*, 389–391.

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.*, *56*, 640–662.

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *J. Neurophysiol.*, *75*, 1843–1857.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cereb. Cortex*, *14*, 452–465.

Morgan, M. L., DeAngelis, G. C., & Angelaki, D. E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron, 59*(4), 662–673.

Odkvist, L. M., Schwarz, D. W., Fredrickson, J. M., & Hassler, R. (1974). Projection of the vestibular nerve to the area 3a arm field in the squirrel monkey (Saimiri sciureus). *Exp. Brain Res.*, *21*, 97–105.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annu. Rev. Neurosci.*, *21*, 227–277.

Previc, F. H. (1992). The effects of dynamic visual stimulation on perception and motor control. *J. Vestib. Res.*, *2*, 285–295.

Royden, C. S., Banks, M. S., & Crowell, J. A. (1992). The perception of heading during eye movements. *Nature*, *360*, 583–585.

Schaafsma, S. J., & Duysens, J. (1996). Neurons in the ventral intraparietal area of awake macaque monkey closely resemble neurons in the dorsal part of the medial superior temporal area in their responses to optic flow patterns. *J. Neurophysiol.*, *76*, 4056–4068.

Schlack, A., Hoffmann, K. P., & Bremmer, F. (2002). Interaction of linear vestibular and visual stimulation in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.*, *16*, 1877–1886.

Schwarz, D. W., & Fredrickson, J. M. (1971a). Rhesus monkey vestibular cortex: A bimodal primary projection field. *Science*, *172*, 280–281.

Schwarz, D. W., & Fredrickson, J. M. (1971b). Tactile direction sensitivity of area 2 oral neurons in the rhesus monkey cortex. *Brain Res.*, *27*, 397–401.

Siegel, R. M., & Read, H. L. (1997). Analysis of optic flow in the monkey parietal area 7a. *Cereb. Cortex*, *7*, 327–346.

Smith, S. T., Bush, G. A., & Stone, L. S. (2002). Amplitude response of human vestibular heading estimation. *Soc. Neurosci. [Abstracts]*, *56*, 1.

Stone, L. S., & Perrone, J. A. (1997). Human heading estimation during visually simulated curvilinear motion. *Vis. Res.*, *37*, 573–590.

Takahashi, K., Gu, Y., May, P. J., Newlands, S. D., DeAngelis, G. C., & Angelaki, D. E. (2007). Multimodal coding of three-dimensional rotation and translation in area MSTd: Comparison of visual and vestibular selectivity. *J. Neurosci.*, *27*, 9742–9756.

Tanaka, K., Fukada, Y., & Saito, H. A. (1989). Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.*, *62*, 642–656.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y., & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *J. Neurosci.*, *6*, 134–144.

Telford, L., Howard, I. P., & Ohmi, M. (1995). Heading judgments during active and passive self-motion. *Exp. Brain Res.*, *104*, 502–510.

van Beers, R. J., Sittig, A. C., & Gon, J. J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *J. Neurophysiol.*, *81*, 1355–1364.

van den Berg, A. V., & Brenner, E. (1994). Why two eyes are better than one for judgements of heading. *Nature*, *371*, 700–702.

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *J. Neurophysiol.*, *76*, 1246–1266.

Warren, W. H. (2003). Optic flow. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1247–1259). Cambridge, MA: MIT Press.

Warren, W. H., Jr., & Hannon, D. J. (1990). Eye movements and optical flow. *J. Opt. Soc. Am. [A]*, *7*, 160–169.

Zhang, T., & Britten, K. H. (2003). Microstimulation of area VIP biases heading perception in monkeys. *Soc. Neurosci. [Abstracts]*, *339*, 9.

35 Visual Stability during Saccadic Eye Movements

CONCETTA MORRONE AND DAVID BURR

ABSTRACT We frequently reposition our gaze by making rapid ballistic eye movements called *saccades* to position the fovea on objects of interest. While the strategy is highly efficient for the visual system, allowing it to analyze the whole visual field with the high resolution of the fovea, it poses several problems for perception. Saccades cause rapid, large-field motion on the retina, potentially confusable with large-field motion in the external world. They also change the relationship between external space and retina position, confounding information about visual direction. Much effort has been made in recent years to attempt to understand the effects of saccades on visual function. Electrophysiological, imaging, and psychophysical evidence suggests that saccades trigger two distinct neural processes: a suppression of visual sensitivity, specific to motion analysis, probably mediated by the magnocellular pathway, and a gross perceptual distortion of visual space just before the repositioning of gaze. While our knowledge of how the visual system copes with the potentially damaging effects of continual saccadic eye movements has increased considerably over the past few decades, many interesting avenues of research remain open.

Vision is always clear and stable, despite continual saccadic eye movements, that is, ballistic movements of the eyes that reposition our gaze two to three times a second. Saccades may be made deliberately, but normally they are automatic and pass unnoticed. An observer at a sporting event, someone conversing with a companion, or a person reading a book usually makes many saccades without knowing that they have occurred. Not only does the actual movement of the eyes escape notice; so too does the motion of images as they sweep across the retina and the fact that gaze itself has been repositioned. The world seems to stay put. Comparable image motion that is produced externally, rather than by movements of the observer's own eyes, has an alarming effect on the observer's sense of stability. The problem of visual stability is an old one that has fascinated many scientists, including Descartes, von Helmholtz, Mach, and Sherrington, and indeed goes back at least to the 11th-century Persian scholar Alhazen: "For if the eye moves in front of visible objects while they are being contemplated, the form of every one of the objects facing the eye . . . will move on the eyes as the latter moves. But sight has become accustomed to the motion of the objects' forms on its surface when the objects are stationary, and therefore does not judge the objects to be in motion" (Alhazen, 1083). But only recently have the tools become available to monitor eye movements accurately and to measure their effects qualitatively.

The problem of visual stability can be broadly divided into three separate issues: Why we do not perceive the *motion* of the retinal image produced as the eye sweeps over the visual field? How do we cope dynamically "on-line" with the continual changes in the retinal image produced by each saccade? How (and where) do we construct a stable *spatiotopic* representation of the world centered in real-world external coordinates from the successive "snapshots" of each fixation? Although the problem of visual stability is far from solved, tantalizing progress has been made over the last few years, some of which will be highlighted in this chapter.

Saccadic suppression

Part of the general problem of visual stability is why the fast motion of the retinal image generated the movement of the eyes completely escapes notice. Comparable wide-field motion generated externally is highly visible and somewhat disturbing (Burr, Holt, Johnstone, & Ross, 1982). It has long been suspected that vision is somehow suppressed during saccades (Holt, 1903), but the nature of the suppression has remained elusive. Now it is clear that the suppression is neither a "central anaesthesia" of the visual system (Holt, 1903), nor a "gray-out of the world" due to fast motion (Campbell & Wurtz, 1978; Dodge, 1900; Woodworth, 1906), as this motion is actually visible—extremely so at low spatial frequencies (Burr & Ross, 1982). What happens is that some stimuli are actively suppressed by saccades while others are not. Stimuli of low spatial frequencies are very difficult to detect if flashed just prior to a saccade, while stimuli of high spatial frequencies remain equally visible (Burr et al., 1982; Volkmann, Riggs, White, & Moore, 1978). *Equiluminant* stimuli (varying in color but not luminance) are not suppressed during saccades and can even be enhanced (Burr,

CONCETTA MORRONE Department of Physiological Sciences, University of Pisa, and Scientific Institute Stella Maris, Pisa, Italy
DAVID BURR Department of Psychology, University of Florence, Italy; School of Psychology, University of Western Australia, Perth, Australia

Morrone, & Ross, 1994), implying that the parvocellular pathway, essential for chromatic discrimination, is left unimpaired, while the magnocellular pathway is specifically suppressed.

Saccadic suppression follows a specific and very tight time course, illustrated in figure 35.1*A* (replotted from Diamond, Ross, & Morrone, 2000). Sensitivity for seeing low-spatial-frequency, luminance-modulated stimuli declines 25 ms before saccadic onset, reaching a minimum at the onset of the saccade, then rapidly recovering to normal levels 50 ms afterward. Does the suppression result from a central nonvisual "corollary discharge" signal (discussed in the next section), or could it result simply from the visual "masking" effects? This would seem unlikely, as great care was taken to ensure a uniform surround. However, the question is important, so to be certain that the saccade itself was essential for the suppression, we simulated saccadic eye movements by viewing the stimulus setup through a mirror that could be rotated at saccadic speeds. When the background was uniform, with minimal visual referents, the simulated saccades had little or no effect on sensitivity (open symbols of figure 35.1*A*).

But that is not to say that masking does not occur under more natural conditions. When the test stimulus is embedded within a textured screen, simulated saccades do decrease contrast sensitivity (figure 35.1*B*). Indeed, the maximum suppression is nearly as great as that caused by real saccades and lasts much longer. This suggests that after the saccade, sensitivity is greater than that expected with comparable motion without the saccade, possibly implying a post-saccadic facilitation (consistent with the physiology).

That real saccades cause a different pattern of results from simulated saccades shows that suppression results at least in part from an active, extraretinal signal. Interestingly, the amount of suppression varies with age, being much stronger in adolescent children than in adults (Bruno, Brambati, Perani, & Morrone, 2006), even though in adolescence, motion perception and masking are largely adultlike (Maurer, Lewis, & Mondloch, 2005; Parrish, Giaschi, Boden, & Dougherty, 2005). This indicates that the mechanisms that mediate suppression are still developing at this age. Because the saccadic motor system is also not completely mature during adolescence (Fischer, Biscaldi, & Gezeck, 1997), this is further evidence that the extraretinal signal that is responsible for mediating the saccadic suppression may be linked to the motor system.

Psychophysical studies indicate that saccadic suppression occurs early in the visual system (Burr et al., 1994), at or before the site of contrast masking and before low-level

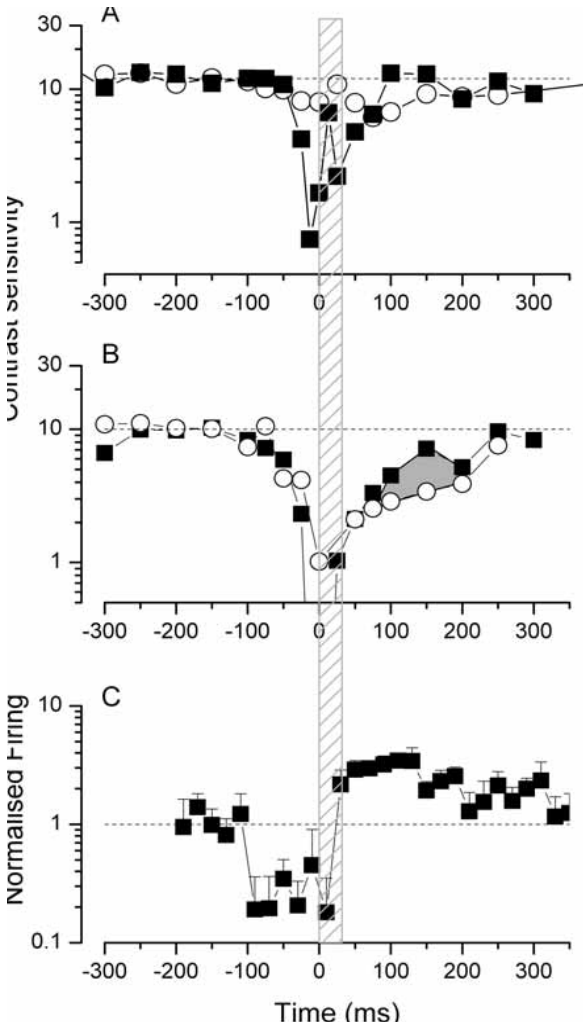


FIGURE 35.1 The effect of saccades on human contrast sensitivity and firing rate in monkey MT. (*A*) Solid squares show contrast sensitivity for discriminating (in two-alternative forced choice) the brightness of a brief, low-frequency, luminance-modulated grating patch as a function of time relative to saccadic onset. The background was of mean luminance, with very few visual referents present. Sensitivity is severely reduced (by more than a log-unit) at saccadic onset. The open circles show measurements made in identical conditions, but instead of making a saccade, a mirror moved at the same speed and amplitude as the saccade; this had very little effect on sensitivity. (*B*) As for panel *A*, except that the background was a high-contrast random check pattern. With a structured background, the simulated saccade did reduce visibility, presumably by masking, with the effect lasting longer than it did for a real saccade. The gray-shaded area indicates the region where sensitivity was greater during the saccade than in fixation. (*C*) Firing rate of a typical MT neuron in an awake monkey in response to stimulation to a brief stimulus, as a function of time relative to saccadic onset. The pattern of the response is similar to the psychophysical results of panel *A*; the timing does not match exactly, but this is only one cell, not an average, and does not take neural latencies into account. The enhancement after the saccade may allow for the more rapid recovery from masking during real rather than simulated saccades (the difference between the solid and open symbols in panel *B*). (Reproduced with permission from Diamond et al., 2000; Ibbotson et al., 2008.)

motion processing (Burr, Morgan, & Morrone, 1999). Thilo, Santoro, Walsh, and Blakemore (2003) addressed this question more directly with a clever electrophysiological technique. Replicating an old study by Riggs, Merton, and Morton (1974), they showed that visual phosphenes produced by electrical stimulation of the eye are suppressed during saccades. But phosphenes of cortical origin—V1 or V2—generated with the technique of transcranial magnetic stimulation were not suppressed. This strongly suggests that saccadic suppression occurs early, before the site of generation of cortical phosphenes, probably within the lateral geniculate nucleus (LGN) or perhaps within V1 itself. A recent functional resonance magnetic imaging (fMRI) (Sylvester, Haynes, & Rees, 2005) study that measured blood oxygen level–dependent (BOLD) activity of LGN while subjects made saccades over a field of constant illumination (to avoid the generation of spurious retinal motion) showed a clear suppression in both LGN and V1, reinforcing early suggestion of saccadic suppression in the dark in V1 (Bodis-Wollner, Bucher, & Seelos, 1999). There is also fMRI evidence for postthalamic modulation by saccades. BOLD activity to luminance stimuli is relatively suppressed compared with that of chromatic stimuli during saccades, but the attenuation varies between areas (Kleiser, Seitz, & Krekelberg, 2004), being strong in MT—as expected—but also strong in V4, a cortical area that receives more parvocellular than magnocellular input. There is also evidence of suppression in higher neural levels, in areas that are normally associated with attention (Bristow, Haynes, Sylvester, Frith, & Rees, 2005, Kleiser et al., 2004). This is interesting, as it could be the suppression of the high-order "attention-related" areas that prevents the sense of motion entering into awareness, causing startle.

The electrophysiology of saccadic suppression is more complex. Electrophysiological studies show that the majority of cells in V1 respond vigorously to the movement created by saccades; however, some cells do not respond to saccade-created motion but only to real motion in the external world. These cells are the minority, about 10% in V1, 15% in V2, and 40% in V3A (Galletti & Fattori, 2003; Wurtz, 2008). Recently, Reppas, Usrey, and Reid (2002) have shown that voluntary saccades induce profound changes in the response of LGN cells, particularly magno-cells. Activity is depressed around the time of the saccade, and there is also a larger and long-lasting enhancement after the saccade. There is also clear evidence for a strong suppression in the colliculus and pulvinar that could be important for the suppression of fast motion (Wurtz, 2008).

Perhaps the data that can be most readily compared with the psychophysical sensitivities are those of Ibbotson, Crowder, Cloherty, Price, and Mustari (2008), who measured responsiveness of MT/MST cells to a brief stimulus, like that used in the psychophysics experiments. An example cell is replotted in figure 35.1*C*. This cell showed a very strong and robust suppression before the start of the saccade, followed by a clear enhancement lasting some 200 ms after the termination of the saccade. While it is difficult to make a quantitative comparison between psychophysical threshold measurements (figure 35.1*A*) and the firing rate of one representative MT cell (figure 35.1*C*), it is interesting that modulation of MT/MST response follows a time course similar to that of sensitivity for a brief low-spatial frequency stimulus, presumably detected by magnocellular/MT-MST pathways. The very strong postsaccadic enhancement of MT cells could explain the relatively higher sensitivity after real saccades compared with after simulated saccades (the difference between open and solid symbols of figure 35.1*B*). Another interesting result reported for MT neurons is that in addition to being suppressed, many neurons seem to reverse their preferred direction selectivity (Thiele, Henning, Kubischik, & Hoffmann, 2002). This odd behavior could be important in "canceling" motion information, helping to keep the world still.

To conclude, it is not surprising that saccadic suppression should occur at different levels. Many basic sensory phenomena, such as gain control, occur not at a single site but at virtually every possible location: photoreceptors, retinal ganglion cells, LGN cells, and cortex (Shapley & Enroth-Cugell, 1984). Indeed, the parallels between saccadic suppression and contrast gain control are strong, suggesting that they might share similar mechanisms. During saccades, the temporal impulse response to luminance, but not to equiluminant, stimuli becomes faster and more transient (Burr & Morrone, 1996). LGN (Reppas et al., 2002) and MT/MST (Ibbotson et al., 2008) cells show a similar response pattern, with faster and more transient impulse response functions during saccades. These results suggest that saccadic suppression might act by attenuating the contrast gain of the neuronal response, causing a faster impulse response (Shapley & Victor, 1981). Changing contrast gain makes neurons less responsive to low-contrast stimuli, decreasing the effectiveness of the spurious noise caused by the saccade, hence facilitating their recovery to normal sensitivity. The fact that saccadic suppression operates via gain control mechanisms is consistent with the fact that the M-pathway is selectively suppressed, as M cells have much stronger gain control mechanisms than P cells (Sclar, Maunsell, & Lennie, 1990). This would certainly be an elegant and economical solution to the problem of saccadic suppression, taking advantage of mechanisms that are already in place for other functions. The idea that gain control explains both the suppression and rapid recovery during saccades has been implemented in a model that simulates quantitatively the time course of contrast sensitivity in normal and simulated saccade (Diamond et al., 2000). In this view, saccadic suppression subserves two important roles: the suppression of image motion, which

would otherwise be disturbing, and the rapid return to normal sensitivity after the saccade.

Dynamic updating of internal spatial maps

Besides the (relatively) simple problem of suppressing the motion caused by the fast-moving image on the retina, the brain must also take into account the saccadic movement when determining the instantaneous position of objects in space. Like Alhazen, von Helmholtz (1866) recognized that "the effort of will involved in trying to alter the adjustment of the eyes" could be used to help stabilize perception. Models based on similar ideas of compensation of eye movements were proposed in the 1950s by Sperry (1950) with the concept of corollary discharge and by Von Holst and Mittelstädt (1954) with the concept efference copy: The effort of will of making the eye movements (corollary discharge) is subtracted from the retinal signal to cancel the eye movement and stabilize perception. Now we know that retinal motion signal cannot be easily compensated, given the sophisticated analysis performed by motion detectors. However, there is evidence for the existence of a corollary discharge signal that must be instrumental in maintaining visual stability.

Considerable psychophysical evidence exists for a corollary discharge in humans, going back to the 1960s, when Leonard Matin and others reported large transient changes in spatial localization at the time of saccades. When asked to report the position of a target that was flashed during a saccade, subjects mislocalized it, primarily in the direction of the saccade (Honda, 1989; Mateeff, 1978; Matin & Pearce, 1965). The localization error is typically on the order of half the saccadic size. Later, Mateeff and Honda measured the time course and showed that the error starts about 50 ms before the saccadic onset and continues well after fixation is regained. The error before the saccadic onset has been taken as an indication of the existence of a slow and sluggish corollary discharge signal that compensates partly for the eye movement; the internal representation of the position and the actual position of the gaze do not match and errors in the localization of a brief target are generated.

We have examined saccadic mislocalization in photopic conditions using equiluminant stimuli (that remain visible during saccades). This approach revealed a bizarre result: At the time of saccades, visual space is not so much shifted in the direction of the saccade but *compressed* toward the saccadic target (Morrone, Ross, & Burr, 1997; Ross, Morrone, & Burr, 1997) (see figure 35.2*A*). Objects that are flashed at saccadic onset to a range of positions, from close to fixation to positions well beyond the saccadic target, are all perceived at or near the saccadic target. The effect is primarily parallel to the saccade direction (Ross et al., 1997), although a small compression is also observed in the orthogonal direction

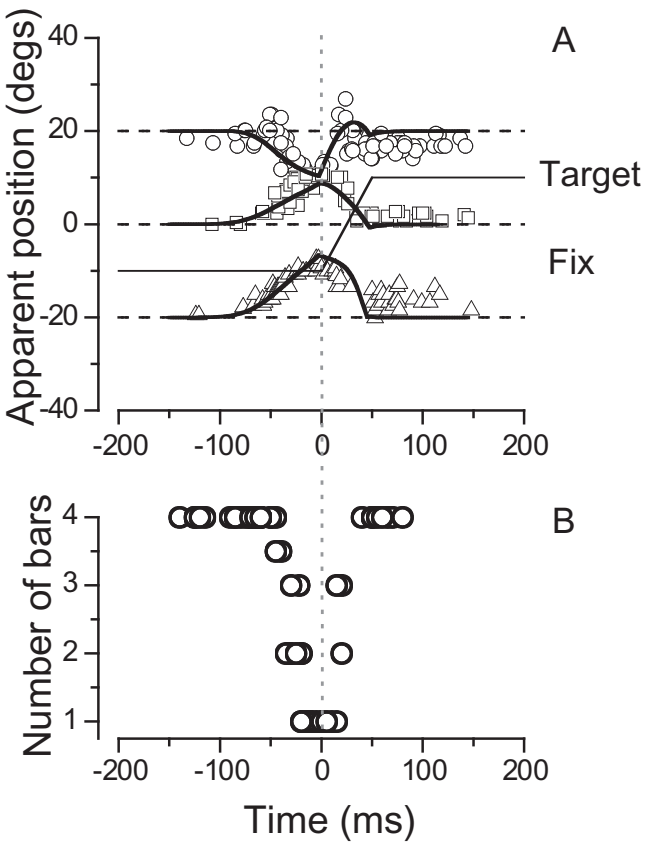


FIGURE 35.2 Effect of saccades on apparent bar position and number. (*A*) Perceived position of narrow green bars, briefly flashed on a red background at various times relative to the onset of a saccade from −10 to +10 degrees. The physical position of the bars (shown by the dashed lines) could be −20 degrees (for the triangle symbols), 0 degrees (square symbols), and 20 degrees (round symbols). The effect of the saccades (maximal at saccadic onset) is to shift the apparent position of the bar toward the saccadic target, where the eyes land. For stimuli at 0 or −20 degrees the shift is in the direction of the saccade, but for stimuli at +20 degrees the shift is in the other direction. In all cases, the shift is toward the saccadic landing point. (*B*) Reported number of bars seen, as a function of presentation time (relative to saccadic onset). A variable number of bars (0, 1, 2, 3, or 4) were presented simultaneously in positions straddling 10 degrees either side of the saccadic target site. The results reported here are for trials in which four bars were presented; but when presented near saccadic onset, the four collapse onto each other, so only one was seen. The other bars were not suppressed, because one bar was always reported as one, and zero bars were reported as zero (no false positives). (Reproduced with permission from Ross et al., 1997.)

(Kaiser & Lappe, 2004). These results are intriguing because they indicate that the process described mathematically by a simple translation of the internal coordinate system is not plausible; perhaps the system cannot perform the transformation of space without additional perceptual costs.

Figure 35.2*B* shows that saccadic compression is so strong that four bars spread over 20 degrees are perceived as being fused into a single bar. Discrimination of shape (Matsumiya

& Uchikawa, 2001) or colors (Lappe, Kuhlmann, Oerke, & Kaiser, 2006) of the bars is still possible, but counting them and perceiving them in separate positions are not. Sometimes the shape or orientation of the flashed object can also change, appearing smaller and more vertical (for horizontal saccades), although these effects have been harder to quantify. The fact that the feature itself is not lost or compressed suggests that the mislocalization occurs at a relatively high level of analysis after feature extraction.

It has been suggested that saccadic compression occurs only when visual references are present and is absent in the dark (Lappe, Awater, & Krekelberg, 2000). However, subsequent studies (Awater & Lappe, 2006) have shown that this is not necessarily true. In the dark or in a transient dark condition achieved by a brief blackout on saccadic onset, compression does occur but can be obscured because in these conditions, there is also a mislocalization of the saccadic target (Morrone, Ma-Wyatt, & Ross, 2005). When this is taken into account, compression occurs in both light and dark, with and without visual references. Several studies have shown that visual references per se (such as scattered points on the monitor) do not affect compression. However, presenting the same brief stimulus twice perisaccadically, even to different retinal locations, greatly reduces mislocalization (Cai, Pouget, Schlag-Rey, & Schlag, 1997; Morrone et al., 1997; Pola, 2007; Sogo & Osaka, 2002), suggesting that the visual system has a mechanism for maintaining positional constancy of objects across saccades.

A related phenomenon is that if the saccadic target is displaced after the saccade has been initiated, the displacement (of up to 30% saccade size) is not noted (Bridgeman, Hendry, & Stark, 1975). However, if there is a brief gap in the reappearance of the target in the displaced position, the displacement is immediately apparent (Deubel, Schneider, & Bridgeman, 1996). This observation led to the idea that the system assumes object stability in the absence of contrary information, probably by comparing presaccadic and postsaccadic positions with some form of short-term memory buffer. These results suggest that the visual system does take advantage of static visual references to help maintain stability across saccades, but the details of how these are selected are stored in some form of memory buffer of limited capacity have yet to be determined.

It has recently been argued that the insensitivity to saccadic target displacement (Bridgeman et al., 1975) can be explained by optimal sensorimotor integration between the retinal signal and extraretinal corollary discharge signals (Niemeier, Crawford, & Tweed, 2003). At the time of saccades, spatial information about eye position, which is necessary to localize objects in external space, is unreliable. Therefore spatial information during this period is given less weight than is information before and after the saccade. The transient distortions of the kind shown in figure 35.2*A* may also be consistent with statistically optimal, or "Bayesian," integration of information. A recent study has shown how this could be the case, by studying audiovisual integration during saccades. Auditory stimuli are usually far more difficult to localize in space than are visual stimuli. When vision and sound are in conflict, vision dominates (the "ventriloquist effect") as predicted by optimal integration. However, when visual stimuli are artificially degraded by blurring, audition can dominate (Alais & Burr, 2004), again consistent with optimal integration. Because saccades have little effect on auditory space perception (Harris & Lieberman, 1996), they are a useful tool to study saccadic mislocalization. Indeed, audiovisual stimuli (bars and beeps presented together in the same spatial position) are mislocalized much less than are visual stimuli that are presented alone, suggesting that visual information is given a low weight during saccades, and this can lead to mislocalization of transient stimuli (Binda, Bruno, Burr, & Morrone, 2007). Not only does the idea explain qualitatively the mislocalization, it explains quantitatively the mislocalization of bimodal audiovisual stimuli over the whole time course relative to saccade onset (figure 35.3).

Binda, Bruno, and colleagues (2007) go on to develop a Bayesian model of saccadic mislocalization, simply assuming, like Niemeier and colleagues (2003), an increase in noisiness of the eye position signal at the time of saccades. At this stage, the model accounts only for the shift in the direction of the saccade, not the accompanying compression. This would require a further assumption, such as a "prior" or "default rule," for objects to be seen at the fovea. While this seems reasonable and has been suggested in other contexts (Deubel, Schneider, & Bridgeman, 2002; MacKay, 1973), it remains speculation at this stage.

So the functional role of spatial compression remains unclear. However, it is interesting that saccadic compression is positively correlated with peak saccadic velocity: Individuals with high saccadic velocity show large compression, while subjects with slow saccadic velocity show mainly a shift in the saccadic direction (but the effect is not related to the spurious visual motion). This suggests a strong link between perception at the time of saccades and the motor system, probably mediated by the corollary discharge signal. It is also interesting to note that the temporal dynamics of saccadic mislocalization are very similar to those of saccadic suppression (compare figures 35.1 and 35.2), indicating a common mechanism, probably the corollary discharge signal. It would be interesting to test whether there is a correlation between saccadic velocity and the magnitude of suppression.

Saccades cause dramatic perceptual localization illusions, but when subjects are required to indicate their response by a motor action—secondary saccades or blind hammering—their responses are near-veridical (Hallett & Lightstone,

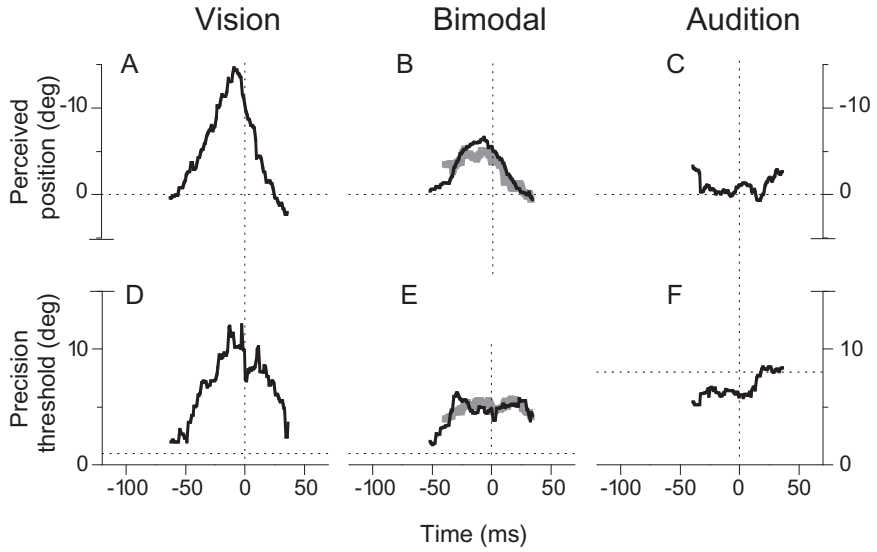


FIGURE 35.3 Illustration of how saccadic mislocalization can result from optimal "Bayesian" fusion. In a two-alternative forced choice, subjects were asked to report whether a perisaccadic test bar that was displayed midway between fixation and saccadic target seemed to be located to the right or left of a presaccadic probe bar. (For full details, see Binda, Bruno, et al., 2007.) Psychometric functions were fitted to these data to give an estimate of perceived position and also of precision of localization. The upper curves show how perceived position varied with time (relative to saccadic onset). Visual stimuli presented on their own (*A*) showed the characteristic mislocalization, like that of figure 35.1*A*. Auditory stimuli were not at all affected by the saccade (*C*). However, when the sound was played contemporaneously with the bar display, the mislocalization of the bar was reduced (*B*). The lower curves show the localization thresholds. Again, sound was unaffected by saccades, but the precision of visual localization was reduced drastically near saccadic onset. During the bimodal audiovisual presentation, precision improved to the extent of being better than either the visual or auditory unimodal localization precision. Indeed, this performance, both for perceived position and for precision thresholds, was very close to the Bayesian prediction, indicated by the thick gray line. The dotted horizontal lines indicate performance during fixation. (Reproduced with permission from Binda, Bruno, et al., 2007.)

1976a, 1976b; Hansen & Skavenski, 1977, 1985). Other studies (e.g., Bridgeman, Lewis, Heit, & Nagle, 1979) also reported that subjects can point accurately to targets that were displaced perisaccadically, even though the subject did not perceive the change in target position. However, a few experiments have failed to replicate the original dissociation between motor accuracy and perceptual error during saccades, reporting localization errors for both tasks (Bockisch & Miller, 1999; Dassonville, Schlag, & Schlag-Rey, 1992, 1995; Honda, 1991; Miller, 1996; Schlag & Schlag-Rey, 1995). Recently, Burr, Morrone, and Ross (2001) and Morrone, Ma-Wyatt, and Ross (2005) reported a clear dissociation between verbal reports and blind pointing for saccadic compression. The plot of figure 35.4 shows that briefly flashed stimuli were perceived clearly in false positions, causing the characteristic compression (solid symbols); but when asked to point blindly at the stimuli, with the screen temporally obscured by liquid crystal shutter, observers did so veridically (open symbols).

Interestingly, analogous effects have been reported in audition. Although saccadic eye movements do not affect the localization of tones, saccadic head movements do (Leung, Alais, & Carlile, 2008). Sounds are compressed toward the endpoint of the head turn. However, if subjects are ask to point to the apparent sound source (by head turn), the compression disappears, as it does for vision (Burr et al., 2001).

However, for visual judgments, introducing clearly visible postsaccadic references under normal lighting conditions causes both verbal report and pointing to show compression. This suggests that vision has access to two maps, one subject to distortion and the other not. The motor map shows no compression except when visual references remain in view for a substantial time after saccade, indicating that these maps are updated postsaccadically, while for perceptual judgments, the updating occurs before and during the actual saccade. Both maps contribute to determining the weight given to each map. Perhaps the popular distinction between conscious perception and action (Goodale & Milner, 1992; Trevarthen, 1968) is at best an oversimplification.

But where in the brain do these maps reside? Is there any evidence that a dynamically updated spatiotopic map actually exists? Electrophysiological studies have reported several transient perisaccadic phenomenon. In the lateral intraparietal cortex (LIP), receptive fields change positional selectivity (Duhamel, Colby, & Goldberg, 1992) just before a monkey makes a saccadic eye movement, anticipating the change in gaze. This is illustrated in figure 35.5*A*, showing the response of an LIP cell to stimuli flashed to the receptive field position

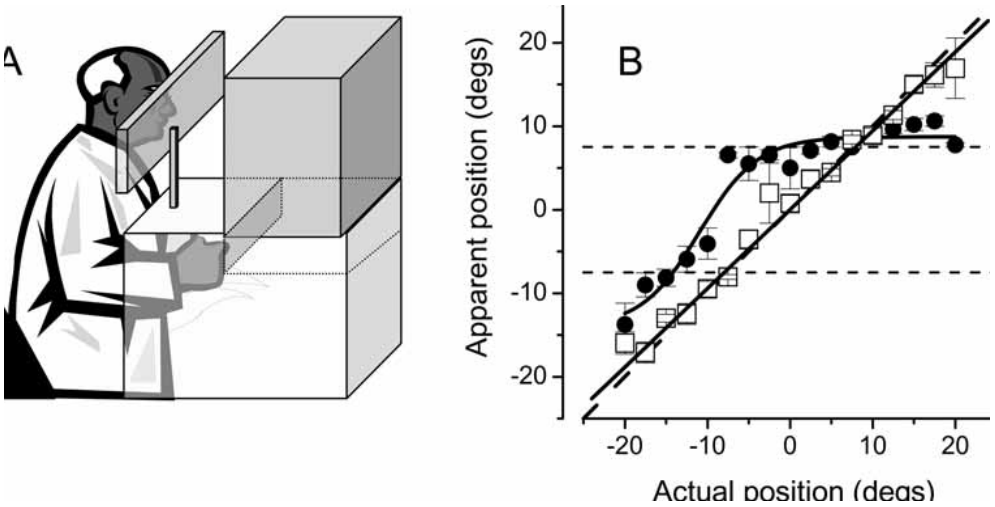


FIGURE 35.4 No spatial compression for motor responses. (*A*) Subjects viewed a cathode-ray tube monitor through a liquid crystal shutter. On command, they made a 15-degree saccade from −7.5 to +7.5 degrees (dashed lines in panel *B*), and a bar was briefly displayed just prior to saccadic onset. Shortly after the saccade was completed, the shutter closed, and subjects responded by jabbing at the touch screen with a brisk ballistic movement, the hand being hidden from view. (*B*) The open squares show the results for the jabbing response for stimuli presented just prior to saccadic onset ($-30 < t < 0$ ms). The responses are near veridical. The solid circles show results for verbal reports, under identical conditions. As shown in figure 35.2, there is a very strong compression, with all stimuli within 10 degrees of the saccadic target seen at saccadic target. (Reproduced with permission from Burr et al., 2001.)

and what will become the receptive field after the saccade has been made ("future receptive field"). Note that the response in the current receptive field starts to reduce and that in the future receptive field starts to increase, long before the eye has actually moved to reposition the retinal image. This is termed *predictive remapping*.

This phenomenon occurs not only in LIP, but also in many other visual areas, including the superior colliculus (Walker, Fitzgibbon, & Goldberg, 1995) and area V3 (Nakamura & Colby, 2002), with area V4 showing a somewhat different behavior (Tolias et al., 2001). It has even been suggested that 10% of neurons in primary visual cortex (V1) show dynamic updating of receptive fields (Nakamura & Colby, 2002). The origin of the phenomenon has been studied in the frontal eye field (FEF), and firm evidence demonstrates that it is mediated by a corollary discharge signal, probably originating in the colliculus and mediodorsal thalamus (Sommer & Wurtz, 2002, 2006). Deactivation of the nucleus abolishes the predictive updating of the receptive field. The corollary discharge signal arrives nearly 100 ms before the updating starts in the FEF, indicating the complexity of the reorganization.

Despite these recent efforts, there are several aspects of the remapping phenomenon that remain unclear. For example, between the time that the neuron starts to respond to stimuli in the updated position and when it regains post-saccadicaly retinotopic specificity, are receptive fields anchored in a transiently craniotopic map? Do the receptive fields undergo changes in size during the remapping? Are the neurons that are susceptible to remapping randomly intermingled with those that do not remap, or is there some specific organization?

Clever psychophysical studies have also demonstrated remapping in humans (Burr & Morrone, 2005; Melcher, 2005, 2007), by studying the spatial selectivity of visual aftereffects. Most aftereffects are spatially selective. But is the selectivity in retinotopic or spatiotopic coordinates? By imposing a saccade between the adaptor and the test, Melcher was able to show that the selectivity was *both* retinotopic and spatiotopic. The degree to which adaptation was spatiotopic varied with the complexity of the aftereffect. Simple adaptation aftereffects, like contrast (thought to be mediated by primary visual cortex) were primarily retinotopic, while more complex aftereffects (such as faces) were primarily spatiotopic; aftereffects of intermediate complexity, like the tilt aftereffect, were both retinotopic and spatiotopic.

Adaptation techniques (Melcher, 2007) can also be used to reveal the dynamics of the updating, by briefly presenting the test just prior to a saccade (figure 35.5B). Long before the saccade, adaptation is maximal when test and adaptor are presented to the same position, at fixation, with very little adaptation at the position of saccadic target. However, when the test is presented perisaccadically but before the eyes have moved, the maximum adaptation occurs for tests near saccadic target, the position that will correspond to the adapted retina after the eyes have moved. The similarity of the time courses of the adaptation and the response of the LIP neuron strongly imply that Melcher's experiment reveals the psychophysical counterpart of the "predictive remapping." At present, it is still uncertain exactly how this transient

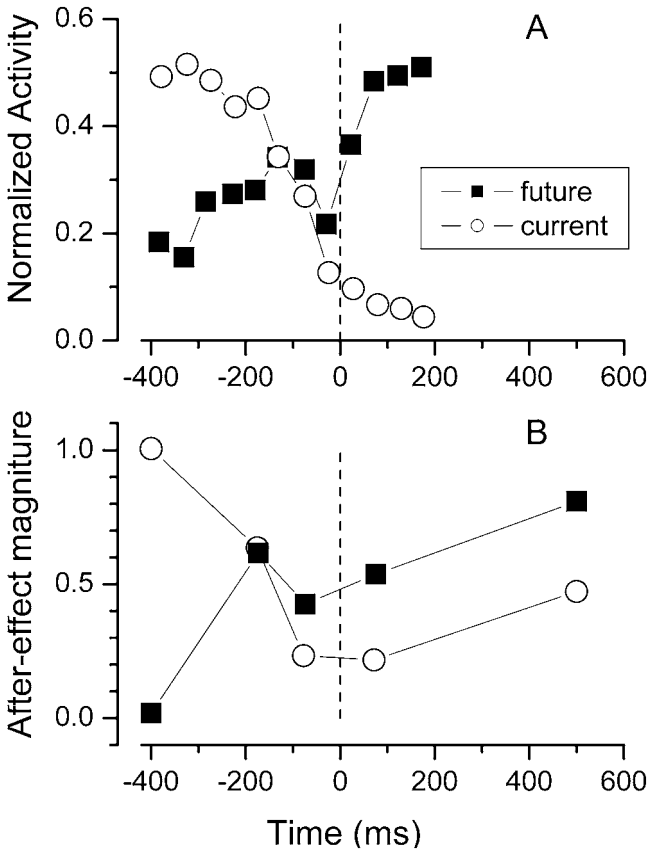


FIGURE 35.5 Predictive remapping in an LIP cell and human observers. (*A*) The response of a "remapping" cell of area LIP of the macaque around the time of the saccade to brief stimuli displayed in the "current" (presaccadic) receptive field (open circles) and to stimuli flashed in what will become its receptive field after the saccade is made. The response to stimuli in the current receptive field begins to decrease before the eyes actually move. Around the same time, the response in the "future" position begins to increase, long before the eyes have actually displaced the receptive field. (*B*) An experiment showing analogous behavior in human psychophysics. Subjects adapted to a tilted grating, then measured the aftereffect to a grating presented in the same (retinal) position ("current," open circles) or to the position that will correspond to the retinal position of the adaptor after a saccade has been made ("future," solid squares). Long before the saccade, there is no adaptation in the future field, and there is full adaptation in the current field (normalized to unity). Like the cell firing rate, adaptation effects in the current field begin to reduce, and those in the future field begin to increase, before the eyes have actually moved. Well after the saccade is terminated, the effects do not drop completely to zero, because this position corresponds to the spatiotopic position of the adaptor, and orientation adaptation has a spatiotopic component (Melcher, 2005). (Reproduced with permission from Kusunoki & Goldberg, 2003, Melcher, 2007.)

updating of receptive fields leads to visual stability, but it is clearly important to test the future activity before the direct input will excite the neuron after the saccade. It could bridge the perception between the two fixations, but can this phenomenon explain perisaccadic mislocalizations?

The dynamics of the remapping receptive field is also very similar to that of perisaccadic mislocalization (figure 35.2), suggesting that a common mechanism could be driving all these phenomena. However, there are several problems in relating the two sets of data quantitatively. Within a framework of labeled-line theory, a neuron that is placed in a specific anatomical position in a cortical map will, when stimulated, signal the presence of a stimulus at that position. However, if it responds presaccadically to a stimulus falling in the *future* receptive field (one displaced in the direction of the saccade), it should still signal this stimulus location as being in the normal location, that is shifted in the direction *opposite* to the saccade: but the results (figures 35.2 and 35.3) show that the primary result is the perception of a shift in the same direction of the saccade. There are two possible schemas to resolve this apparent contradiction. The first is to consider that the remapped activity of the future receptive field is the neuronal response of the corollary discharge signal, mapped in retinal coordinates (Binda, Bruno, et al., 2007). This activity is present only if a visual stimulus is present; it is active only when important information needs to be updated, reducing also the complexity of the phenomenon. Within this framework, the addition (fusion) of this activity with the retinotopic activity of visual cortex could generate the shift of apparent positions in the appropriate direction, as we have recently demonstrated for the audiovisual targets (figure 35.3). The other possibility is to consider that the remapped neuronal activity is not referred to the exact time of the stimulus presentation but is read after the saccade is complete, in a form of *postdiction* (Eagleman & Sejnowski, 2000). This would also imply that perceptual time should be altered by saccades, as indeed it is (see below).

At the time of the saccade, the timing of the neuronal response changes dramatically. In all cells of areas V3A and FEF that remap during saccades, their remapped response is faster than that during fixation (Nakamura & Colby, 2002). Similarly, the latencies of neurons in areas MT and MST are shorter in response to real saccade than to simulated saccades (Price, Ibbotson, Ono, & Mustari, 2005). These effects have psychophysical implications: Saccades cause a compression, and even an inversion, of perceived time (Morrone, Ross, & Burr, 2005). When asked to compare the perceived duration of a temporal interval presented around the time of a saccade with one presented 2 s afterward, subjects judged it much shorter, about half the duration (figure 35.6). Again the time course of this distortion is quite tight and, after taking into account the duration of the stimuli, similar to that of the spatial compression.

Preliminary data (Binda, Burr, & Morrone, 2007) also indicate that the perceived time at saccadic onset, measured using an auditory tone, is delayed about 100 ms, while about 50 ms before saccade, the latency is reduced by about 20 ms, consistent with the inversion of time data and with

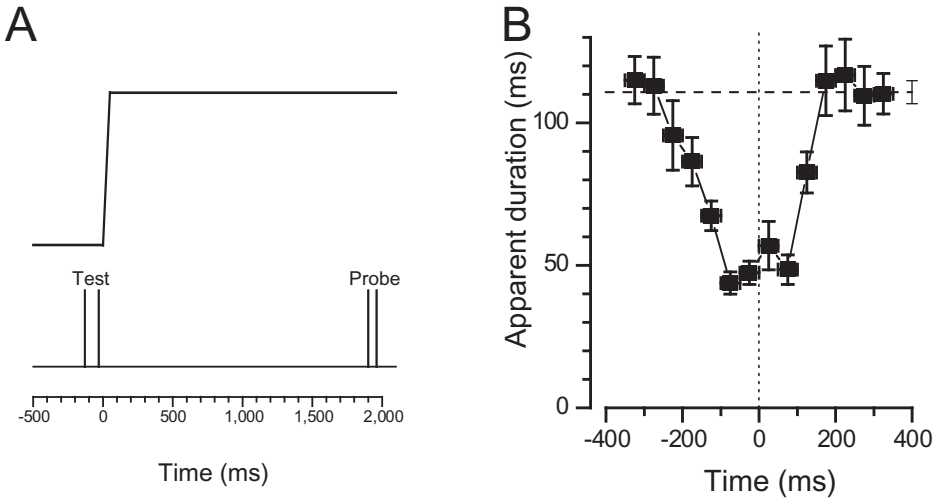


FIGURE 35.6 Time is also compressed during saccades. (*A*) The subject was asked to compare the duration of the interval of two test flashes (separated by 100 ms) with a postsaccadic probe of variable duration that appeared 2 s later. (*B*) The apparent duration was then calculated from psychometric functions. Around the time of saccadic onset, apparent duration was about half the physical duration. The dashed line shows the duration match during fixation. (Reproduced with permission from Morrone, Ross, & Burr, 2005.)

the fact that during the remapping neuronal latency became shorter about 40 ms, explaining the inversion. In addition, they also indicate that stimuli that are presented at saccadic onset are coincident with stimuli presented soon after saccades, facilitating the interpretation of their position in the postsaccadic coordinate system.

Space and time are generally studied separately and thought of as separate and independent dimensions. However, as we have observed, both space and time undergo severe transient distortions at the time of saccades, as objects become compressed toward the saccadic target (Ross et al., 1997), and perceived temporal durations are severely shrunk (Morrone, Ross, and Burr, 2005). As was discussed above, the relationship between perceptual shifts and receptive field updating is far from clear, and compression of time and of space is even more difficult to understand. Nevertheless, we can advance a few firm properties that might help to explain compression. As the transient changes both in space and in time follow very similar dynamics, they might well be manifestations of a common neural cause, a distortion in the space-time metric (Morrone, Ross, & Burr, 2005). Compression of relative distances in space and time are consistent with a reduction of spatial and temporal sampling. This is also one of the few concepts that would explain the perisaccadic increase in sensitivity for size (Santoro, Burr, & Morrone, 2002) and duration (Morrone, Ross, & Burr, 2005) judgments.

If together with the undersampling, the receptive field becomes transiently oriented in space-time such that stimuli presented before the saccade and near the fixation are integrated with stimuli presented later for position far from fixation, we could provide a description of the origin of all perisaccadic phenomena. The rotation in space and time of the neuronal selectivity is a concept that has strong and important analogies to the physical rotation on space and time that occurs in motion at relativistic speeds, discussed in detail elsewhere (Morrone, Ross, & Burr, 2008). Unfortunately, the dynamics of the changes in receptive fields during saccades are not yet well enough described to pursue this idea much further at present.

Transsaccadic integration and craniotopic maps

Our normal experience comes from the information derived from one fixation being transferred to the next, even when a particular object or part of the scene becomes hidden. Theories about transsaccadic integration have abounded over the past decades. Early ideas (e.g., Jonides, Irwin, & Yantis, 1982) assumed a "transsaccadic memory buffer" that accumulated high-precision information from each saccade to construct a detailed representation of the world (like pinning tails on a donkey). These ideas fell out of favor, largely because of the implicit implication that the visual system must construct some form of stable Cartesian theater to be viewed by a homunculus. More recent theories have swung to the opposite extreme, assuming that perceptual stability depends, paradoxically, on the lack of internal representation of the world (O'Regan & Noe, 2001). Observers are largely insensitive to transsaccadic changes in the visual scene, questioning how much detailed visual information can be gleaned by making an eye movement on demand; many have assumed that no visual memory is necessary at all (Findlay & Gilchrist,

2003; McConkie & Zola, 1979; Tatler, 2001). In practice, however, it is still necessary for the brain to know where to look for the information that it needs, since eye movements are not random and are rarely wasted in natural tasks (Land, Mennie, & Rusted, 1999; Najemnik & Geisler, 2003). Thus some information about the layout of the scene and the position of important objects must somehow be represented and accumulated across saccades. There is clear evidence showing that at least three or four objects are transferred successfully across saccades even in the absence of allocentric cues (Prime, Niemeier, & Crawford, 2006).

Recently, Melcher and Morrone (2003) showed that transsaccadic integration occurs for motion signals that are individually below threshold (and hence are not perceived when presented alone). Two periods of coherent horizontal motion (150 ms each) were shown successively, separated by sufficient time to allow for a saccadic eye movement between them. On some blocks of trials, subjects saccaded across the stimulus between the two motion intervals; on others, they maintained fixation above or below the stimulus. Thresholds were similar in the two conditions, showing that the motion signals were temporally integrated across the saccade—but only when the two motion signals were in the same position in space, indicating that the brain must use a mechanism that is anchored to external rather than retinal coordinates. Importantly, the methodology excluded cognitive strategies or verbal recoding, since the motion signals presented before and after the saccade were each well below the conscious detection threshold; only by summating the two signals could motion be correctly discriminated.

Another example of craniotopic mechanisms is the demonstration of spatially specific adaptation of event-time (Burr, Tozzi, & Morrone, 2007), showing that adaptation to a fast-moving (20 Hz) spatially localized grating decreases the apparent duration of gratings that are presented to that part of the visual field (in external space) but not to other spatial locations.

Because of the spatial selectivity of individual neurons, the response of primary and secondary visual cortex forms a *map* (Morgan, 2003), similar in principle to that imaged on the retinae (except for distortions due to magnification of central vision). This *retinotopic* representation, which changes completely each time the eyes move, forms the input for all further representations in the brain. So a major question is how this retinotopic representation becomes transformed into the *spatiotopic* representation that we perceive, anchored in stable real-world coordinates.

Electrophysiological studies have shown that neurons in specific areas of associative visual cortex, including V6 (Galletti, Battaglini, & Fattori, 1993) and VIP (Duhamel, Bremmer, BenHamed, & Graf, 1997), do show the spatiotopic selectivity that we would expect to exist; their tuning is invariant of gaze, unlike areas V1 and V2 (that provide their input). Unfortunately, the exact transformation from retinal to spatiotopic coordinates is not yet fully understood, although the suggestion has been made that Bayesian fusion of the retinal signal with eye position signals is sufficient in principle to generate spatiotopic maps, probably acting via eye position–dependent modulation of the neural response, also referred to as *gain fields* (Pouget, Deneve, & Duhamel, 2002; Snyder, Grieve, Brotchie, & Andersen, 1998; Zipser & Andersen, 1988).

Functional magnetic resonance imaging has also indicated the existence of spatiotopic coding in human cortex, both in LO (McKyton & Zohary, 2006), an area deputed to the analysis of objects and in MT+ (d'Avossa et al., 2007; Goossens, Dukelow, Menon, Vilis, & van den Berg, 2006). Using stimuli similar to those used by Melcher and Morrone, our group has reported that the response of a portion of human MT complex varies with gaze position in a way that is consistent with spatiotopic coding. The results are illustrated in figure 35.7. With gaze fixed in the centre of the screen, both areas V1 and MT show spatial selectivity, responding only when the stimuli are presented to the contralateral field (figures 35.7*A* and 35.7*B*). However, if the stimulus is fixed (in the center) and its retinal projection varied by varying gaze, the results are different. V1 still responds only to the contralateral stimulus, but MT responds to both ipsilateral and contralateral stimuli, equally strongly. Further experiments suggested that MT actually shifts its receptive fields to cause spatiotopic coding.

However, it must be pointed out that this result is currently controversial, and contrary results have been reported. Gardner, Merriman, Movshon, and Heeger (2008) report that under the conditions of their experiment, the response of MT is retinotopic rather than spatiotopic. One interesting difference between the two studies is that in Gardner and colleagues' experiment (but not in d'Avossa and colleagues' experiment) attention was directed toward the fovea. We have recently replicated the conditions of their experiment and shown that when attention is withdrawn from the stimulus, the spatiotopic mapping changes to a retinotopic mapping (Crespi et al., 2009). Why attention should be necessary for the remapping is far from clear, but this suggests the operation of normalizing gain control. Fully understanding this mechanism will be an interesting future challenge.

The fact that spatiotopic (or at least craniotopic) coding is more common in the dorsal area might suggest that it could be used for the action system. As was discussed above, the action system seems to update spatial maps much later than the perceptual system does. Perhaps the updating of craniotopic maps takes time but leads to more robust coding of information, explaining the resistance of this system to saccadic mislocalization. The perceptual system, on the other hand, might operate not with a complete map anchored in external coordinates but with ensembles of neurons with

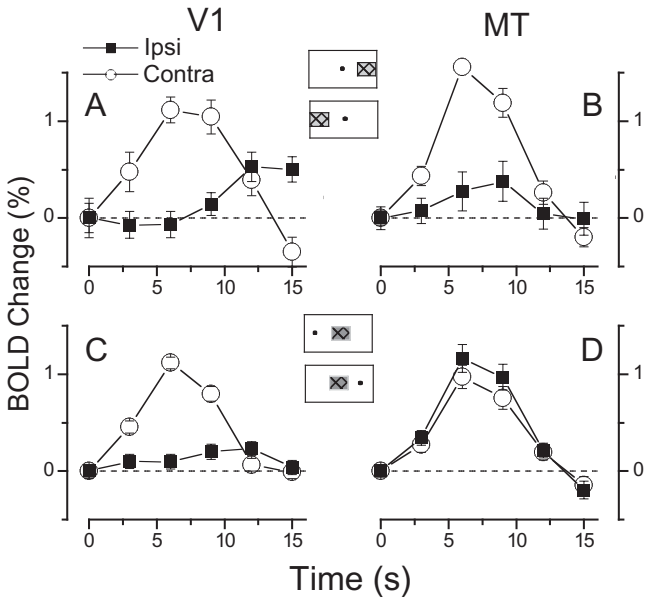


FIGURE 35.7 The dependence of MT response on gaze. (*A*, *B*) BOLD responses for areas V1 (*A*) and MT (*B*) to ipsilateral (solid squares) and contralateral (open circles) stimulation, keeping gaze fixed at screen center (see icons). In all conditions, subjects were required to discriminate the direction of motion to keep sustained attention on the stimulus. As expected, both areas respond strongly only to contralateral stimulation. (*C*, *D*) Response of V1 (*C*) and MT (*D*) to ipsilateral and contralateral stimulation varying gaze; the stimulus was in screen center, and the subject fixated to the left or right of it (see icon). The response pattern of V1 was unchanged, suggesting that V1 is selective only to retinal position, irrespective of gaze. However, the response pattern of MT was quite different, now being strong for both ipsilateral and contralateral stimulation, suggesting that it is selective to the position of the stimulus on the screen, not on the retina. (Reproduced with permission from d'Avossa et al., 2007.)

receptive fields anchored in retinotopic coordinates but transiently shifted just before each saccade (like the neuron of figure 35.5). This transient updating on each saccade might be sufficient to maintain a useful perceptual representation that is not actually in spatiotopic coordinates that take longer to develop.

The important conclusion from these and related studies is that the visual system does combine information from one fixation to the next but that this process is not like sticking postage stamps on a tailor's dummy: detailed "snapshots" are not integrated within a transsaccadic buffer that preserves the external metric (Jonides et al., 1982). Indeed, such a scheme could be problematic, as scenes do change continuously as objects move and rotate. Inappropriate *pixelwise* integration could lead to very weird percepts, like cubist art. Transsaccadic integration does not occur at the pixel level, but after a certain amount of visual processing, so attributes such as form, orientation, motion, and even complex entities such as faces are integrated across fixations. This in itself does not solve the problem of visual stability, but it could provide a basis for visual continuity with ever-changing retinal input. It might well be that the two processes—dynamic receptive field updating and craniotopic coding—collaborate in the selection of the important information to be integrated. The remapping neurons are primed before the eye movement actually occurs, so they can determine whether the information of successive fixations should be integrated. Perhaps if activation during remapping were constant, then craniotopic receptive fields could receive a switch signal allowing information to be integrated transsaccadicaly. If the remapping neurons do not respond, the switch could open, vetoing the integration of craniotopic receptive fields, so they accumulate new information starting afresh. Within this schema, both the integration across saccades and perisaccadic mislocalization might involve the same mechanisms to obtain a stable vision across separate glances without fusion of local, pixel-like visual details.

Conclusion

Seeing is usually believing. For about two-thirds of our waking lives, we perceive objects where vision tells us they are, which, more often than not, coincides with their actual position. In the remaining time, the visual system sends us erroneous spatial information, presumably because it is engaged in correcting the troublesome consequences of eye movements on retinal afferences. When this happens, we disbelieve visual information. If available, spatial cues from other senses become dominant; if we have to act, we use the robust representation of the craniotopic system without attempting to update it dynamically. If vision is the only signal that is present, we deform our concept of space and of time to make sense of it and to not miss visual information for more than one-third of our waking time.

ACKNOWLEDGMENTS This work was supported by European Union PF6 NEXT (MEMORY) and PF7 IDEAS: STANIB.

REFERENCES

ALAIS, D., & BURR, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.*, *14*(3), 257–262.

ALHAZEN, I. (1083). Book of optics. In A. I. Sabra (Ed.), *The optics of Ibn al-Haytham*. London: Warburg Institute, 1989.

AWATER, H., & LAPPE, M. (2006). Mislocalization of perceived saccade target position induced by perisaccadic visual stimulation. *J. Neurosci.*, *26*(1), 12–20.

BINDA, P., BRUNO, A., BURR, D. C., & MORRONE, M. C. (2007). Fusion of visual and auditory stimuli during saccades: A Bayesian explanation for perisaccadic distortions. *J. Neurosci.*, *27*(32), 8525–8532.

BINDA, P., BURR, D. C., & MORRONE, M. C. (2007). Spatiotemporal distortions of visual perception during saccades. *Perception, 36* (ECVP Abstract Supplement), 112.

Bockisch, C., & Miller, J. (1999). Different motor systems use similar damped extraretinal eye position information. *Vis. Res.*, *39*, 1025–1038.

Bodis-Wollner, I., Bucher, S. F., & Seelos, K. C. (1999). Cortical activation patterns during voluntary blinks and voluntary saccades. *Neurology*, *53*, 1800–1805.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of visual world during saccadic eye movements. *Vis. Res.*, *15*, 719–722.

Bridgeman, B., Lewis, S., Heit, G., & Nagle, M. (1979). Relation between cognitive and motor-oriented systems of visual position perception. *J. Exp. Psychol. Hum. Percept. Perform.*, *5*(4), 692–700.

Bristow, D., Haynes, J. D., Sylvester, R., Frith, C. D., & Rees, G. (2005). Blinking suppresses the neural response to unchanging retinal stimulation. *Curr. Biol.*, *15*(14), 1296–1300.

Bruno, A., Brambati, S. M., Perani, D., & Morrone, M. C. (2006). Development of saccadic suppression in children. *J. Neurophysiol.*, *96*(3), 1011–1017.

Burr, D., & Morrone, M. C. (2005). Eye movements: Building a stable world from glance to glance. *Curr. Biol.*, *15*(20), R839–R840.

Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nat. Neurosci.*, *10*(4), 423–425.

Burr, D. C., Holt, J., Johnstone, J. R., & Ross, J. (1982). Selective depression of motion sensitivity during saccades. *J. Physiol.*, *333*, 1–15.

Burr, D. C., Morgan, M. J., & Morrone, M. C. (1999). Saccadic suppression precedes visual motion analysis. *Curr. Biol.*, *9*, 1207–1209.

Burr, D. C., & Morrone, M. C. (1996). Temporal impulse response functions for luminance and colour during saccades. *Vis. Res.*, *36*, 2069–2078.

Burr, D. C., Morrone, M. C., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature*, *371*, 511–513.

Burr, D. C., Morrone, M. C., & Ross, J. (2001). Separate visual representations for perception and action revealed by saccadic eye movements. *Curr. Biol.*, *11*(10), 798–802.

Burr, D. C., & Ross, J. (1982). Contrast sensitivity at high velocities. *Vis. Res.*, *23*, 3567–3569.

Cai, R. H., Pouget, A., Schlag-Rey, M., & Schlag, J. (1997). Perceived geometrical relationships affected by eye-movement signals. *Nature*, *386*, 601–604.

Campbell, F. W., & Wurtz, R. H. (1978). Saccadic ommission: Why we do not see a greyout during a saccadic eye movement. *Vis. Res.*, *18*, 1297–1303.

Crespi, S., Biagi, L., Burr, D. C., d'Avossa, G., Tosetti, M., & Morrone, M. C. (2009). Spatial attention modulates the spatiotopicity of human MT complex. *Perception, 38* (ECVP Abstract Supplement).

d'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nat. Neurosci.*, *10*(2), 249–255.

Dassonville, P., Schlag, J., & Schlag-Rey, M. (1992). Oculomotor localization relies on a damped representation of saccadic eye movement displacement in human and nonhuman primates. *Vis. Neurosci.*, *9*, 261–269.

Dassonville, P., Schlag, J., & Schlag-Rey, M. (1995). The use of egocentric and exocentric location cues in saccadic programming. *Vis. Res.*, *35*, 2191–2199.

Deubel, H., Schneider, W. X., & Bridgeman, B. (1996). Postsaccadic target blanking prevents saccadic suppression of image displacement. *Vis. Res.*, *36*, 985–996.

Deubel, H., Schneider, W. X., & Bridgeman, B. (2002). Transsaccadic memory of position and form. *Prog. Brain Res.*, *140*, 165–180.

Diamond, M. R., Ross, J., & Morrone, M. C. (2000). Extraretinal control of saccadic suppression. *J. Neurosci.*, *20*, 3442–3448.

Dodge, R. (1900). Visual perception during eye movements. *Psychol. Rev.*, *7*, 454–465.

Duhamel, J., Bremmer, F., BenHamed, S., & Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature*, *389*, 845–848.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, *255*(5040), 90–92.

Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science*, *287*(5460), 2036–2038.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing.* Oxford, UK: Oxford University Press.

Fischer, B., Biscaldi, M., & Gezeck, S. (1997). On the development of voluntary and reflexive components in human saccade generation. *Brain Res.*, *754*(1–2), 285–297.

Galletti, C., Battaglini, P. P., & Fattori, P. (1993). Parietal neurons encoding spatial locations in craniotopic coordinates. *Exp. Brain Res.*, *96*, 221–229.

Galletti, C., & Fattori, P. (2003). Neuronal mechanisms for detection of motion in the field of view. *Neuropsychologia*, *41*(13), 1717–1727.

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *J. Neurosci.*, *28*(15), 3988–3999.

Goodale, M. A., & Milner, A. D. (1992). Separate pathways for perception and action. *Trends Neurosci.*, *15*, 20–25.

Goossens, J., Dukelow, S. P., Menon, R. S., Vilis, T., & van den Berg, A. V. (2006). Representation of head-centric flow in the human motion complex. *J. Neurosci.*, *26*(21), 5616–5627.

Hallett, P. E., & Lightstone, A. D. (1976a). Saccadic eye movements towards stimuli triggered by prior saccades. *Vis. Res.*, *16*(1), 99–106.

Hallett, P. E., & Lightstone, D. (1976b). Saccadic eye movements to flashed targets. *Vis. Res.*, *16*, 107–114.

Hansen, R. M., & Skavenski, A. A. (1977). Accuracy of eye position information for motor control. *Vis. Res.*, *17*(8), 919–926.

Hansen, R. M., & Skavenski, A. A. (1985). Accuracy of spatial locations near the time of saccadic eye movments. *Vis. Res.*, *25*, 1077–1082.

Harris, L. R., & Lieberman, L. (1996). Auditory stimulus detection is not suppressed during saccadic eye movements. *Perception*, *25*(8), 999–1004.

Holt, E. B. (1903). Eye movements and central anaesthesia. *Psychol. Rev.*, *4*, 3–45.

Honda, H. (1989). Perceptual localization of visual stimuli flashed during saccades. *Percept. Psychophys.*, *46*, 162–174.

Honda, H. (1991). The time courses of visual mislocalization and of extra-retinal eye position signals at the time of vertical saccades. *Vis. Res.*, *31*, 1915–1921.

Ibbotson, M., Crowder, N., Cloherty, S., Price, N., & Mustari, M. (2008). Saccadic modulation of neural responses: Possible roles in saccadic suppression, enhancement, and time compression. *J. Neurosci.*, *28*, 10952–10960.

Jonides, J., Irwin, D. E., & Yantis, S. (1982). Integrating visual information from successive fixations. *Science*, *215*(4529), 192–194.

Kaiser, M., & Lappe, M. (2004). Perisaccadic mislocalization orthogonal to saccade direction. *Neuron*, *41*(2), 293–300.

Kleiser, R., Seitz, R. J., & Krekelberg, B. (2004). Neural correlates of saccadic suppression in humans. *Curr. Biol.*, *14*(5), 386–390.

Kusunoki, M., & Goldberg, M. E. (2003). The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. *J. Neurophysiol.*, *89*(3), 1519–1527.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Lappe, M., Awater, H., & Krekelberg, B. (2000). Postsaccadic visual references generate presaccadic compression of space. *Nature*, *403*, 892–895.

Lappe, M., Kuhlmann, S., Oerke, B., & Kaiser, M. (2006). The fate of object features during perisaccadic mislocalization. *J. Vis.*, *6*(11), 1282–1293.

Leung, J., Alais, D., & Carlile, S. (2008). Compression of auditory space during rapid head turns. *Proc. Natl. Acad. Sci. USA*, *105*, 6492–6497.

MacKay, D. M. (1973). Visual stability and voluntary eye movements. In R. Jung (Ed.), *Handbook of sensory physiology* (Vol. VII/3, pp. 307–331). Berlin: Springer-Verlag.

Mateeff, S. (1978). Saccadic eye movements and localization of visual stimuli. *Percept. Psychophys.*, *24*(3), 215–224.

Matin, L., & Pearce, D. G. (1965). Visual perception of direction for stimuli flashed during voluntary saccadic eye movements. *Science*, *148*, 1485–1487.

Matsumiya, K., & Uchikawa, K. (2001). Apparent size of an object remains uncompressed during presaccadic compression of visual space. *Vis. Res.*, *41*(23), 3039–3050.

Maurer, D., Lewis, T. L., & Mondloch, C. J. (2005). Missing sights: Consequences for visual cognitive development. *Trends Cogn. Sci.*, *9*(3), 144–151.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across succesive fixations in reading? *Percept. Psychophys.*, *25*, 221–224.

McKyton, A., & Zohary, E. (2006). Beyond retinotopic mapping: The spatial representation of objects in the human lateral occipital complex. *Cereb. Cortex*, *17*, 1164–1172.

Melcher, D. (2005). Spatiotopic transfer of visual-form adaptation across saccadic eye movements. *Curr. Biol.*, *15*(19), 1745–1748.

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nat. Neurosci.*, *10*(7), 903–907.

Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nat. Neurosci.*, *6*(8), 877–881.

Miller, J. (1996). Egocentric localization of a perisaccadic flash by manual pointing. *Vis. Res.*, *36*, 837–851.

Morgan, M. J. (2003). *The space between your ears: How the brain represents visual space*. London: Weidenfeld & Nicolson.

Morrone, M. C., Ma-Wyatt, A., & Ross, J. (2005). Seeing and ballistic pointing at perisaccadic targets. *J. Vis.*, *5*(9), 741–754.

Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nat. Neurosci.*, *8*(7), 950–954.

Morrone, M. C., Ross, J., & Burr, D. C. (1997). Apparent position of visual targets during real and simulated saccadic eye movements. *J. Neurosci.*, *17*, 7941–7953.

Morrone, M. C., Ross, J., & Burr, D. C. (2008). Keeping vision stable: Rapid updating of spatiotopic receptive fields may cause relativistic-like effects. In R. Nijhawan (Ed.), *Problems of space and time in perception and action*. Cambridge, UK: Cambridge University Press.

Najemnik, J., & Geisler, W. S. (2003). Optimal visual search. *J. Vis.*, *3*(9), 624.

Nakamura, K., & Colby, C. L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proc. Natl. Acad. Sci. USA*, *99*(6), 4026–4031.

Niemeier, M., Crawford, J. D., & Tweed, D. B. (2003). Optimal transsaccadic integration explains distorted spatial perception. *Nature*, *422*(6927), 76–80.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.*, *24*(5), 939–973; discussion 973–1031.

Parrish, E. E., Giaschi, D. E., Boden, C., & Dougherty, R. (2005). The maturation of form and motion perception in school age children. *Vis. Res.*, *45*(7), 827–837.

Pola, J. (2007). A model of the mechanism for the perceived location of a single flash and two successive flashes presented around the time of a saccade. *Vis. Res.*, *47*(21), 2798–2813.

Pouget, A., Deneve, S., & Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nat. Rev. Neurosci.*, *3*(9), 741–747.

Price, N. S., Ibbotson, M. R., Ono, S., & Mustari, M. J. (2005). Rapid processing of retinal slip during saccades in macaque area MT. *J. Neurophysiol.*, *94*(1), 235–246.

Prime, S. L., Niemeier, M., & Crawford, J. D. (2006). Transsaccadic integration of visual features in a line intersection task. *Exp. Brain Res.*, *169*(4), 532–548.

Reppas, J. B., Usrey, W. M., & Reid, R. C. (2002). Saccadic eye movements modulate visual responses in the lateral geniculate nucleus. *Neuron*, *35*(5), 961–974.

Riggs, L. A., Merton, P. A., & Morton, H. B. (1974). Suppression of visual phosphenes during saccadic eye movements. *Vis. Res.*, *14*, 997–1011.

Ross, J., Morrone, M. C., & Burr, D. C. (1997). Compression of visual space before saccades. *Nature*, *384*, 598–601.

Santoro, L., Burr, D., & Morrone, M. C. (2002). Saccadic compression can improve detection of Glass patterns. *Vis. Res.*, *42*(11), 1361–1366.

Schlag, J., & Schlag-Rey, M. (1995). Illusory localization of stimuli flashed in the dark before saccades. *Vis. Res.*, *35*, 2347–2357.

Sclar, G., Maunsell, J. H., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vis. Res.*, *30*(1), 1–10.

Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. In J. G. Osborn & N. N. Chadler (Eds.), *Progress in retinal research* (Vol. 3, pp. 263–346). Oxford, UK: Pergamon Press.

Shapley, R. M., & Victor, J. D. (1981). How the contrast gain control modifies the frequency responses of cat retinal ganglion cells. *J. Physiol. Lond.*, *318*, 161–179.

Snyder, L. H., Grieve, K. L., Brotchie, P., & Andersen, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature*, *394*(6696), 887–891.

Sogo, H., & Osaka, N. (2002). Effects of inter-stimulus interval on perceived locations of successively flashed perisaccadic stimuli. *Vis. Res.*, *42*(7), 899–908.

Sommer, M. A., & Wurtz, R. H. (2002). A pathway in primate brain for internal monitoring of movements. *Science*, *296*(5572), 1480–1482.

Sommer, M. A., & Wurtz, R. H. (2006). Influence of the thalamus on spatial visual processing in frontal cortex. *Nature*, *444*(7117), 374–377.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.*, *43*, 482–489.

Sylvester, R., Haynes, J. D., & Rees, G. (2005). Saccades differentially modulate human LGN and V1 responses in the presence and absence of visual stimulation. *Curr. Biol.*, *15*(1), 37–41.

Tatler, B. W. (2001). Characterising the visual buffer: Real-world evidence for overwriting early in each fixation. *Perception*, *30*(8), 993–1006.

Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K. P. (2002). Neural mechanisms of saccadic suppression. *Science*, *295*(5564), 2460–2462.

Thilo, K. V., Santoro, L., Walsh, V., & Blakemore, C. (2003). The site of saccadic suppression. *Nat. Neurosci.*, *7*, 13–14.

Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., & Schiller, P. H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, *29*(3), 757–767.

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychol. Forsch.*, *31*, 299–348.

Volkmann, F. C., Riggs, L. A., White, K. D., & Moore, R. K. (1978). Contrast sensitivity during saccadic eye movements. *Vis. Res.*, *18*, 1193–1199.

von Helmholtz, H. (1866). *Handbuch der Physiologischen Optik*. (Reprinted in J. P. C. Southall (Ed.), *A treatise on physiological optics*. New York: Dover, 1963.)

von Holst, E., & Mittelstädt, H. (1954). Das Reafferenzprinzip. *Naturwissenschaften*, *37*, 464–476.

Walker, M. F., Fitzgibbon, J., & Goldberg, M. E. (1995). Neurons of the monkey superior colliculus predict the visual result of impending saccadic eye movements. *J. Neurophysiol.*, *73*, 1988–2003.

Woodworth, R. S. (1906). Vision and localization during eye movements. *Psychol. Bull.*, *3*, 68–70.

Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vis. Res.*, *48*(20), 2070–2089.

Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, *331*(6158), 679–684.

36 Optimal Estimation in Sensory Systems

EERO P. SIMONCELLI

ABSTRACT A variety of experimental studies suggest that sensory systems are capable of performing estimation or decision tasks at near-optimal levels. In this chapter, I explore the use of optimal estimation in describing sensory computations in the brain. I define what is meant by optimality and provide three quite different methods of obtaining an optimal estimator, each based on different assumptions about the nature of the information that is available to constrain the problem. I then discuss how biological systems might go about computing (and learning to compute) optimal estimates.

The brain is awash in sensory signals. How does it interpret these signals so as to extract meaningful and consistent information about the environment? Many tasks require estimation of environmental parameters, and there is substantial evidence that the system is capable of representing and extracting very precise estimates of these parameters. This is particularly impressive when one considers that the brain is built from a large number of low-energy, unreliable components, whose responses are affected by many extraneous factors (e.g., temperature, hydration, blood glucose and oxygen levels).

The problem of optimal estimation has been well studied in the statistics and engineering communities, in which a plethora of tools have been developed for designing, implementing, calibrating, and testing such systems. In recent years, many of these tools have been used to provide benchmarks or models for biological perception. Specifically, the development of signal detection theory led to widespread use of statistical decision theory as a framework for assessing performance in perceptual experiments. More recently, optimal estimation theory (in particular, Bayesian estimation) has been used as a framework for describing human performance in perceptual tasks.

In this chapter, I will explore the use of optimal estimation in describing sensory computations in the brain. In the first half, I will define what I mean by optimality and will develop three quite different formulations for obtaining an optimal estimator. In the second half, I will ask how biological systems might go about computing optimal estimates. This is not intended as a complete review of this rich multidisciplinary topic, and I apologize in advance to the many authors whose important contributions I have neglected to mention. Instead, my purpose is to clarify and resolve a number of myths and misunderstandings about optimal estimation and to offer a personal perspective on the relationship between these concepts and the design and function of biological sensory systems.

EERO P. SIMONCELLI Center for Neural Science and Courant Institute of Mathematical Sciences, New York University, New York, New York

Definition and formulations of optimal estimation

A common problem for systems that must interact with the world (including both biological organisms and human-made devices) is that of obtaining estimates of environmental properties, x, from sensory measurements, m. An *estimator* is simply a deterministic function, $f(m)$, that maps measurements to values of the variable of interest. If x is a binary variable, then the estimator reduces to a decision function. Generally, the measurements are assumed to be corrupted by noise, which could arise from a number of sources, including the signal itself (e.g., the quantization of light into photons, when one is interested in knowing the light intensity), the transduction mechanism, or variability within the neurons that are transmitting and computing with this information (see Faisal, Selen, & Wolpert, 2008), for a recent review of noise in the nervous system).

Our primary question is: How does an organism select and implement a good estimator or (more optimistically) the best estimator? To address this, we will have to state explicitly what we mean by *best*. The traditional statistical formulation of the best estimator is the one that minimizes the average value of a predefined loss (cost) function, $L(x, f(m))$. The loss function specifies the cost of generating an estimated value of $f(m)$ when the true value is x. It is generally assumed to be positive and equal to zero only when the estimate is equal to the true value.

REGRESSION FORMULATION Suppose we wanted to build a machine that could perform optimal estimation of x, given a noisy measurement m.[1] We can imagine "training" this

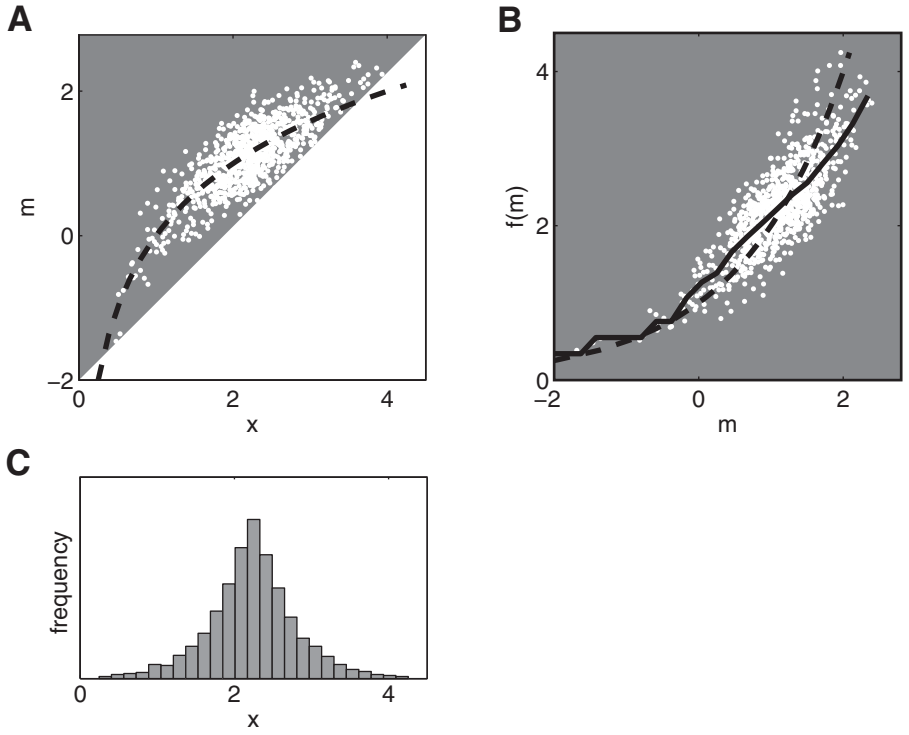


FIGURE 36.1 Regression formulation of the optimal estimation problem, illustrated for a one-dimensional signal and measurement. (*A*) The measurement process (also known as the *encoding* process). We assume a set of data pairs (plotted points), $\{x_n, m_n\}$, indexed by $n \in [1, 2, \ldots, N]$, representing true signal values and associated noisy measurements. The dashed line indicates the average measurement as a function of the true signal value. (*B*) The estimation (or *decoding*) process. The estimator $f(m)$ maps measurements back to estimated signal values. The optimal estimator (solid line) does this so as to minimize a specified loss function. Note that this need not be (and is generally not) the inverse of the average measurement function (dashed line). Note also that the optimal estimator will depend on the signal values that are included in the data set, which are summarized by the histogram shown in panel *C*. (See color plate 49.)

machine by showing it many signal-measurement pairs, $\{x_n, m_n\}$. Typically, we imagine that each measurement arises from its associated true value through some sort of noisy transformation. Figure 36.1*A* illustrates such a set of training data.

An estimator, f, attempts to invert the measurement process, mapping measurements m back to signal values x (figure 36.1*B*). This mapping is deterministic: Each measurement leads to a unique estimate. But if we hold the signal value fixed and make a set of estimates (each arising from a different measurement), these estimates will fluctuate because of the variability in the underlying measurements. The optimal estimator is the one that minimizes the average loss over these examples:

$$f_{\text{opt}} = \arg\min_f \frac{1}{N} \sum_n L(x_n, f(m_n))$$

We will refer to this as a "regression" estimator; a special case is the linear regression solution, which arises when L is the squared error. An example optimal estimator is indicated by the solid line in figure 36.1*B*. Note that this transformation is not the same as the inverse of the transformation to average measurements (i.e., the inverse of the dashed line shown in figure 36.1*A*).

Of course, the precision with which we can constrain the function f depends on how much data we have. Loosely speaking, the usual approach is to restrict f to be sufficiently simple (e.g., smooth, or defined by a small number of parameters) that the available data will constrain it properly. For example, the estimator shown in figure 36.1*B* was computed by binning the data (as a function of m) and computing the best estimate value for each bin. More formally, we might specify a restricted set of possible functions (denoted $\mathcal{F}$) from which the solution will be selected.[2] Finally, note that the solution we obtain will depend on the *distribution* of data. If the set of training examples includes many x values clustered in a particular region of the space, then the average loss will contain many terms from that region, and the optimization process will thus attempt to reduce the estimation errors there, typically at the expense of larger errors elsewhere. This suggests that the training examples should be selected to represent the distribution of values that might be encountered in the environment.

The regression formulation is appealing because it is simple and intuitive. Its primary limitation is that it requires *supervised* training. That is, obtaining an optimal estimator

relies on a training set of noisy measurements, m_n, each accompanied by its corresponding correct signal value, x_n. Supervised learning for estimation and classification problems has been well studied. A standard example is the problem of learning an input-output relationship with a simplified network of artificial neurons, for which the optimal solution may be obtained by backpropagation (essentially, a form of stochastic gradient descent on the objective function). But this requires large amounts of data, especially when learning multidimensional functions.

From the biological/behavioral perspective, a fully supervised training paradigm also seems implausible. Although most organisms absorb enormous amount of sensory data during their lifetimes, the information that they receive regarding "correct" answers would seem to be relatively sparse. For example, consider the problem of estimating the distance to a nearby object on the basis of visual input. We can compare our estimate to the one that is obtained by reaching out and touching the object. But the amount of this kind of feedback we receive seems vastly insufficient to train the enormous cascade of neurons that are involved in estimating distances from visual input. Similarly, optimization through natural selection (with surviving organisms passing preferred solutions to their offspring genetically) seems implausible, both because of the time required and because genetic material seems unlikely to contain sufficient information to encode even a fraction of the detailed connectivity of those neurons.

Instead, it seems that evolution has endowed the brain with powerful capabilities for *unsupervised* learning (based on noisy measurements alone) and that this is used to supplement and bolster the supervised learning that may be used in the relatively infrequent cases for which the correct answers are known. Unsupervised learning is a heavily studied topic in machine learning (e.g., Hinton & Sejnowski, 1999), and methods have been developed for learning patterns in data, mostly for purposes of optimal coding or clustering/categorization. Perhaps less well known is the fact that optimal estimators may also be written in unsupervised form. To explain this, I will turn first to a probabilistic formulation of the problem.

Probabilistic (Bayesian) Formulation When we describe optimality in terms of minimizing an objective function over a training data set, we usually have in mind that this set is representative of future data we will encounter. This notion may be formalized by describing both the training and future data as samples randomly drawn from a common probability distribution. The law of large numbers tells us that as the number of data pairs grows, the original regression objective function will converge to the expected value (mean) of the loss function, integrated over all possible combinations of x and m:

$$f_{\text{opt}} = \arg\min_f \iint P(x, m)\, L(x, f(m))\, dx dm$$

Unlike the regression formulation, which is written directly in terms of data, the probabilistic formulation is written in terms of a continuous probability density. Since this formulation effectively results from assuming infinite amounts of data, the smoothness constraint that was necessary for selecting an estimator in the regression case is now optional.

The probabilistic objective function may be simplified by using the definition of conditional probability to rewrite the joint density as a product of the marginal density of m and the conditional density of x given m (known as the *posterior* distribution):

$$f_{\text{opt}} = \arg\min_f \int P(m) \int P(x|m)\, L(x, f(m))\, dx dm$$

If the estimator is unrestricted, then we may ignore the outer integral and optimize the estimator separately for each measurement value:

$$f_{\text{opt}}(m) = \arg\min_f \int P(x|m)\, L(x, f(m))\, dx$$

That is, for each measurement, the best estimate is the one that minimizes the expected value of the loss function over the posterior distribution for that measurement.

Finally, the posterior distribution may be rewritten in terms of densities that are more naturally associated with the process from which the data arise. Specifically, we can the describe measurement noise using a conditional probability $P(m|x)$. This *measurement density* expresses the probability of m for each value x of the signal. If we think of it the other way around, holding the measurement fixed and reading off a function of the signal, this is known as a *likelihood function*. Now Bayes' rule can be used to express the posterior in terms of the measurement density and the *prior* distribution $P(x)$, which expresses the probability of occurrence of value x in the world:

$$f_{\text{opt}}(m) = \arg\min_f \int \frac{P(m|x)P(x)}{P(m)} L(x, f(m))\, dx \qquad (1)$$

An example of the Bayesian solution, based on the same distributions that were used to generate the data in figure 36.1, is illustrated in figure 36.2. To provide some intuition, it is worth mentioning several well-known special cases.

Quadratic error (least squares) solution The most common case used in the engineering community is the *least squares* loss function, $L(x, f(m)) = (x - f(m))^2$. In this case, the optimal estimate (which can be derived by differentiating the objective function and setting equal to zero) is simply the mean of the posterior:

$$f_{\text{LS}}(m) = \int x P(x|m)\, dx$$

It is worth mentioning that in the special case of a jointly Gaussian probability density over signal and measurement,

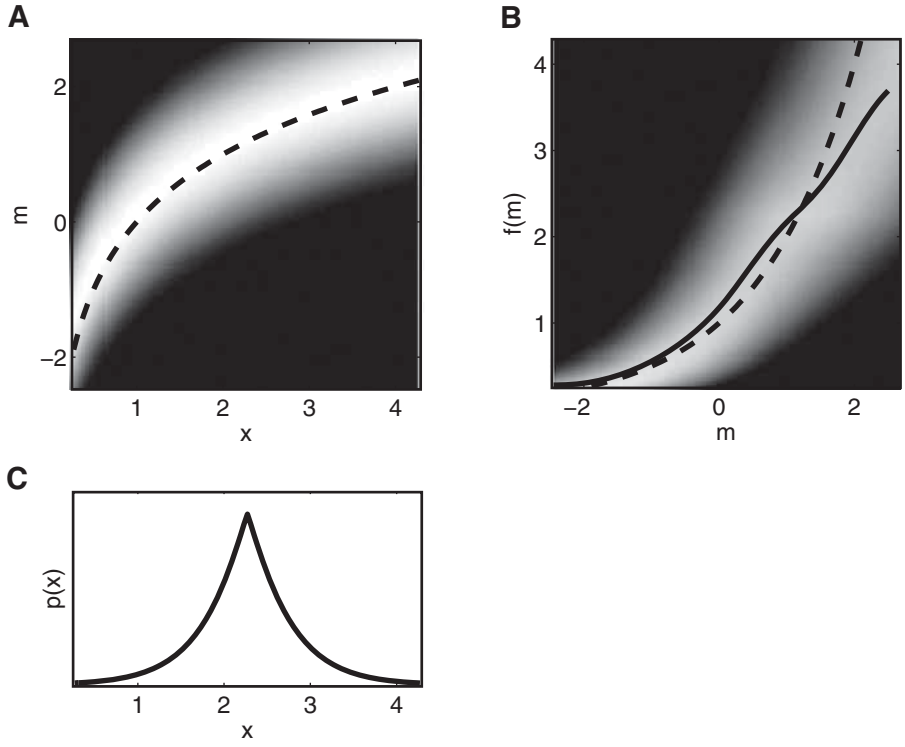


FIGURE 36.2 Bayesian formulation of the optimal estimation problem. (*A*) The measurement density, $P(m|x)$, shown as a gray-scale image, where intensity indicates log probability. The dashed line indicates the mean of the density as a function of x. (*B*) The posterior density, $P(x|m)$. The solid line indicates the mean of the density, and the dashed line indicates the (inverted) mean of the measurement density in panel *A*. (*C*) the prior density, $P(x)$. (See color plate 50.)

this solution turns out to be a linear function of the measurement (the solution is the same as that of our next example).

Linear estimator, quadratic error Now consider what happens when the estimator is restricted to be a linear function of the measurement. The *linear least squares* solution is

$$f_{\mathrm{LLS}}(m) = \frac{\sigma_{xm}}{\sigma_{mm}} m$$

The linear solution relies only on the cross-correlation between signal and measurement and between the measurement and itself, and not on full knowledge of the posterior density. This result extends naturally to multidimensional inputs or outputs.

Maximum probability solution Suppose that the loss function penalizes all errors equally except for the correct answer (which incurs no penalty). Then the solution is the maximum of the posterior density, known as the *maximum a posteriori* (MAP) estimator:

$$f_{\mathrm{MAP}}(m) = \arg\max_x P(x|m)$$

In summary, the probabilistic formulation expresses the estimation problem in terms of four natural ingredients:

- The prior, $P(x)$, which represents the probability of encountering different signal values in the world
- The measurement density, $P(m|x)$, which represents the (probabilistic) relationship between the signal and measurement
- The loss function, $L(x, f(m))$, which represents the cost of making errors
- The family of functions $\mathcal{F}$ from which the estimator is to be chosen. (This ingredient might not be required for the Bayesian solution, which effectively operates under conditions of infinite data.)

Note that although the regression solution of the previous seciton was developed directly from pairs of input-output data, it is also implicitly relying on these same ingredients. Specifically, it is effectively based on the joint probability density of signal and measurement, which is equivalent to the product of the prior likelihood. And as was stated in the previous section, it also requires the specification of a loss function and a family of functions from which the solution is to be drawn.

It is worth emphasizing the most obvious implication of this ingredient list, since it is often misunderstood. Optimality is not a fixed universal property of an estimator but one that depends on each of these defining ingredients; statements about optimality that do not fully specify the ingredients are therefore relying on hidden assumptions. For example, many authors assume that optimality implies that

an estimator must be unbiased (that is, on average, computes the correct value). But many well-known optimal estimators exhibit bias (in fact, all of the estimator examples mentioned above can exhibit bias, depending on the specific choices of prior and measurement densities).[3]

Despite the appealing decomposition of the problem into intuitively sensible ingredients, the probabilistic formulation has drawbacks. In particular, the reliance of the regression solution on supervised training data has been replaced with reliance on knowledge of two abstract probability densities. Since the measurement density is a property of the sensory system, we might imagine learning it from a set of calibration measurements or assuming that it is a fixed property of the device. On the other hand, the Bayesian formulation is often criticized for reliance on the unknown (and perhaps unknowable) prior distribution,[4] and this criticism is further inflamed by the many examples in the literature that introduce a prior as an ad hoc function that may be freely chosen to make the solution tractable. In the view set out here, the prior is meant to capture the statistical structure of some aspect of the world, an assumption that is only slightly stronger than the assumption that the training data in the regression estimator are representative of future data that the system will need to process.

Unsupervised Learning of Optimal Estimators The Bayesian view assumes that all ingredients of the problem, including the prior, are known. If the prior is meant to correctly represent the distribution of signal values in the world, it must presumably be learned from measurements. Engineers who need to design real systems generally follow one of several practical solutions: (1) Directly measure the distribution of signal values that might be encountered by the device, and use a model of this empirical distribution for the prior; (2) assume a prior distribution of some parametric form, and then adjust the parameters so as to best explain the observed distribution of noisy measurements; or (3) assume an estimator of some parametric form and adjust this directly to improve performance on observed data. The first solution requires a separate set of uncorrupted signal measurements and therefore does not seem relevant to biological systems, which are presumably able to make measurements only through the same noisy sensors from which they will be making their estimates. The second solution is generally known as *empirical Bayesian* estimation, since the prior is obtained from noisy training data. The third solution, as described, is simply the regression solution, which relies on supervised training data in order to measure and optimize the estimator performance. Remarkably, it can sometimes be rewritten in an unsupervised form. Below, I will consider the second and third solutions in more detail.

The empirical Bayes formulation assumes a prior of a known parametric form,[5] optimizes the parameters to fit the (noisy) measurements, and then uses this optimized prior to obtain the estimator (by minimizing equation 1). The prior parameters are typically chosen to make the observed data as consistent as possible with the model, and this is usually achieved in practice by maximizing the probability of the observed data:

$$\theta_{\text{opt}} = \arg\max_{\theta} \prod_{n} \int P_{\theta}(x)\, P(m_n|x)\, dx$$

Beyond the potential difficulties associated with computing this optimization, the introduction of this probabilistic cost function is a bit inconsistent, since the prior that best explains the data is not necessarily the one that will lead to the best estimator. Nevertheless, empirical Bayes solutions are often quite successful in situations in which the data are sufficient to strongly constrain the prior parameters.

The third solution mentioned above can, in some cases, be obtained from unsupervised training (Raphan & Simoncelli, 2007). As such, we will refer to it as *unsupervised regression*. The derivation of the general form is somewhat complex, but the simplest case (due to Stein, 1981) arises in the context of an additive Gaussian noise model and a squared-error loss function and can be written quite simply. Stein showed that the mean squared error in this case can be rewritten in a form that depends only on the measurements and not the signal:

$$\iint P(x,m)[x - f(m)]^2\, dx\, dm = \int P(m)\left[g(m)^2 + 2\sigma^2 g'(m) + \sigma^2\right] dm,$$

where $g(m) = f(m) - m$, and σ is the standard deviation of the additive Gaussian noise (assumed to be known). This remarkable result implies that the squared error may be approximated by averaging over measured (noisy) data without knowledge of the correct answers (i.e., unsupervised) and with no assumption about the prior, $P(x)$. This implies that we can select an optimal estimator f (or, equivalently, g) by minimizing the integrand above, averaged over a set of noisy measurement data. As with the original regression solution, the estimator must be restricted sufficiently (e.g., drawn from some parametric family) that it can be constrained by the available data. Analogous expressions can be derived for a number of other measurement probabilities, (Raphan & Simoncelli, 2007).

In both the empirical Bayesian and unsupervised regression formulations, we have exchanged the supervised data pairs required by the standard regression solution for unsupervised data and a known (or previously calibrated) description of the measurement density, $P(m|x)$. We can thus view these solutions as a compromise between the data-oriented regression form and the more abstract Bayesian form. A summary of the ingredients that are required by each of the optimal estimators introduced thus far is provided in table 36.1.

TABLE 36.1

Ingredients required for specifying/learning various formulations of optimal estimator

	Measurement Values $\{m_n\}$	Signal Values $\{x_n\}$	Measurement Probability $P(m \mid x)$	Signal Probability $P(x)$	Loss Function L	Estimator Family $\mathcal{F}$
Regression	✓	✓			✓	Parametric
Bayesian			✓	✓	✓	
Empirical Bayesian	✓		✓	Parametric	✓	
Unsupervised (e.g., Stein) regression	✓		Restricted		Quadratic	Parametric

Ingredients required for specifying/learning various formulations of optimal estimator (see text for definitions). Checkmarks indicate that the ingredient is required but unrestricted. Unlabeled spaces indicate that the ingredient is not needed. Note that the unsupervised regression estimator has been derived only for certain specific measurement densities.

Based on Raphan & Simoncelli (2007).

Optimal estimation in the brain

In this section, we ask how the optimal estimation formulations developed in the previous section can be used in modeling biological sensory systems and how these models can be tested experimentally. These questions can be addressed at many levels, and in this short chapter, I will not attempt to provide a complete overview. Rather, I will describe few published results and try to explain what I see as some of the more important challenges that we currently face in this endeavor.

The concept that sensory perception arises through the fusion of incoming sensor measurements with one's prior experience is often attributed to Hermann von Helmholtz (1925). Although his descriptions are qualitative and do not mention noise or loss functions, they do capture the essence of the Bayesian formulation described in the previous section. This interpretation of perception seems to have lain dormant from von Helmholtz's day until the 1950s, when E. T. Jaynes, a statistically minded physicist, submitted an article to *IRE Transactions on Information Theory*, in which he proposed that Bayesian estimation might be used as a framework for modeling sensory transformations (Jaynes, 1957). The journal rejected the article (on the grounds that it was too speculative), and the concept appears to have lain dormant for another 30 years! In the interim, perceptual psychologists began using signal detection theory as a framework for analyzing psychophysical data (Green & Swets, 1966) and for providing an upper bound on performance. This methodology often does not include explicit loss functions and rarely includes a prior, but the formalization nevertheless represents an important step toward the optimal estimation framework.

PERCEPTUAL BAYESIANISM In the 1980s and 1990s, there was a dramatic revival of the Bayesian methodology across many fields, and perceptual science was one of these. A variety of experiments have aimed to test optimality of human estimation judgments by comparing performance to an "ideal observer" model (e.g., Barlow, 1980; Geisler, 1989; Kersten, 1990; Knill, Field, & Kersten, 1990). A number of reviews document the activity to date (Knill & Richards, 1996; Maloney, 2002; Mamassian, Landy, & Maloney, 2002; Kersten, Mamassian, & Yuille, 2004; Körding, 2007), and this endeavor has been expanded by recent activity in "neuroeconomics," a cross-disciplinary enterprise that aims to characterize decision-making and more general behavioral processes with respect to prior probabilities and reward contingencies (Glimcher, Camerer, Poldrack, and Fehr, 2008).

What does it mean to say that a human subject is performing optimally? As I have emphasized in the first part of this chapter, the definition of the word *optimal* requires specification of a set of ingredients: the measurement probability, the prior, the loss function, and (in some cases) a family of estimators. Specifying these ingredients for a human observer performing a particular task is often difficult or impossible. For example, specifying the measurement probability requires knowledge of how the signal of interest is represented within the brain (including a specification of the noise). In some experiments, investigators have incorporated noise into the stimulus, which can provide insights into the properties of internal noise (and thus the measurement probability) (e.g., Pelli & Farell, 1999; Körding & Wolpert, 2004). The specification of an appropriate family of estimators should be determined by the set of computations that can potentially be performed by neurons, but we currently lack a detailed description of this set.

The loss function can pose more substantial difficulties. Subjects may differ inherently in the way they behave in an experimental situation (e.g., consider personality traits such as risk aversion versus thrill-seeking). Even in cases in which the investigator attempts to control for this by building a loss function directly into the design of the experiment (for example, by paying/penalizing subjects for correct/incorrect answers), one does not know a priori whether or how the subject will learn and internalize these costs, what type

of training (e.g., supervised versus unsupervised) this would require, and how long it would take.

Last, consider the prior, for which one can ask the same questions as were asked for the loss function (whether/how a subject internalizes it, what type of training is required, over what time scales). We might imagine that the observer operates according to a prior that was obtained over a relatively slow timescale (much longer than the duration of a typical experiment, say, on a developmental or evolutionary timescale). In this case, the investigator might attempt to measure it from the environment (or derive it from a model of the environment). At the other extreme, we might imagine that the subject's internal prior is quite flexible and that over the duration of the experiment, the subject internalizes the distribution of stimuli that have been presented. In fact, many experiments are designed so that the subject can learn the distribution of signal values over a set of training trials.

In the context of learning a prior, one apparent paradox seems worth mentioning. Adaptation to stimuli that persist for timescales of seconds or minutes has been found, for every sensory modality and for a wide range of stimulus configurations, to produce substantial changes in subsequent perception. For example, adaptation to a visual stimulus, say, of a given orientation, induces biases in the perceived orientation of subsequently viewed stimuli. These biases are generally *repulsive*: The perceived orientation of a postadaptation stimulus is pushed away from that of the adaptor. If we were to interpret the adaptation as a means by which the system updates its prior probability distribution for orientation, we might expect that heavy exposure to the adapting stimulus would cause an increase in the internally represented prior probability of that stimulus, which should then lead to an *attractive* bias in subsequent perception! Thus adaptation over these timescales appears to be inconsistent with learning of prior probabilities. An alternative interpretation, still within the Bayesian framework, is that adaptation effects correspond to a change in the likelihood function (Stocker & Simoncelli, 2006b).

Given the difficulty of specifying the ingredients of an optimal estimator, one can consider an alternative experimental approach for exploring optimal estimation theories of perception. Specifically, one can ask: For what choices of ingredients would the subject's behavior be considered optimal? The trial-averaged estimates of a human subject do not place a sufficient constraint on the problem to answer this question. In particular, the average response of an estimator does not uniquely determine the prior, likelihood, and loss function that could have been used to define it. But if one assumes, say, a quadratic loss function and measures not just the average response but the full distribution of estimates, then it is possible to extract a prior (Paninski, 2006) or both the likelihood and prior (Stocker & Simoncelli, 2006a) from the psychophysical data.

Ultimately, it seems important to move beyond the initial question of *whether* an observer is optimal to examine the prior and cost conditions under which the observer may be considered optimal, and the flexibility of that optimality. Given that the estimators must operate under changing conditions, we'd like to know: (1) which ingredients of the problem may be learned or adjusted, and what type of adjustments can be made (e.g., Körding & Wolpert, 2004), (2) what type of learning is possible (e.g., supervised, unsupervised, direct verbal communication), (3) over what time scales this learning occurs, and (4) whether the observer is able to switch between estimators (or ingredients of estimators, such as the prior) that have been previously learned (e.g., Körding & Wolpert, 2004; Maloney & Mamassian, 2009).

Physiological Implementation In addition to interpreting perception in the context of optimal estimation, we can consider how such optimal computations might be implemented in the brain. The responses of sensory neurons are commonly described in terms of their selectivity to particular parameters of the stimulus. In most cases, no single neuron is responsible for encoding a stimulus parameter. Instead, the parameter is jointly represented by a population of neurons with different tuning properties; therefore any estimate of the parameter requires a combination of information across many cells, if not the whole population. Over the past 20 years, the theoretical neuroscience community has been exploring the means by which neural responses might be optimally "read out" to explain behavior (e.g., Georgopoulos, Schwartz, & Kettner, 1986; Bialek, Rieke, de Ruyter van Steveninck, & Warland, 1991; Anderson & van Essen, 1994; Seung & Sompolinsky, 1993; Potters & Bialek, 1994; Salinas & Abbott, 1994; Sanger, 1996; Snippe, 1996; Shadlen, Britten, Newsome, & Movshon, 1996; Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997; Rieke & Baylor, 1998; Zhang, Ginzburg, McNaughton, & Sejnowski, 1998; Zemel, Dayan, & Pouget, 1998; Platt & Glimcher, 1999; Gold & Shadlen, 2000; Simoncelli, 2003; Pouget, Dayan, & Zemel, 2003; Bialek & van Steveninck, 2005; Jazayeri & Movshon, 2006).

It is worth noting that despite my emphasis on the probabilistic formulation of the problem, computation of optimal estimates does not necessarily require that the brain explicitly represent or compute probabilities. As was described in the previous section, given the four ingredients of the optimal estimation problem, an estimator is just a fixed deterministic function that maps noisy measurements to estimated values. If these ingredients are fixed and unchanging, the brain could learn to compute the optimal estimator using either regression or one of the two unsupervised methods described previously and would not need to explicitly calculate or represent probabilities!

In general, we imagine that the ingredients of the optimal estimation problem *do* change; the loss function is typically task-dependent, the prior may change gradually (or even suddenly, for example, when the observer moves into a different environment), and the measurement probability may also change (owing to physiological changes in the neural substrate). But even under these conditions, the solution need not explicitly require the calculation or representation of probabilities. If the prior and/or measurement probabilities are parametric, then the optimal estimator is just a function that depends on those parameters. In this case, the parameters may be computed from previous measurements, either through supervised regression or by using the empirical Bayesian formulation. Again, this does not require explicit representation of probabilities.

Although explicit probability representation is not required for optimal estimation, much published work assumes it. A simple means of encoding probabilities, suggested by a number of authors (e.g., Hinton, 1992; Foldiak, 1993; Simoncelli, 1993; Anderson & van Essen, 1994; Sanger, 1996; Gold & Shadlen, 2000; Weiss & Fleet, 2002; Eliasmith & Anderson, 2002; Sahani & Dayan, 2003; Simoncelli, 2003; Barber, Clark, & Anderson, 2003), assumes that the firing rates of each neuron directly represents the probability (or the log probability) of a particular stimulus parameter value. In this view, the population encodes (either directly or through a set of linear basis functions) a probability distribution over the parameter. Downstream neurons could then explicitly compute an estimate from this population, for example, by computing the population mean (a weighted sum) or peak (winner-takes-all). For example, suppose that the firing rates in a given neural population represent the posterior probability evaluated at a set of different signal values. The mean of the density can be computed as a weighted sum over the responses (the weights will be determined by both the signal values at which the posterior is sampled and the portion of the signal space covered by each neuron). Or subsequent stages could operate on the posterior information, postponing the explicit determination of an estimate until it is needed. In either case, the prior in this model can be adjusted by changing the gain on each of the neurons.

The explicit representation of uncertainty, through the breadth and shape of population responses, and the possibility of linear readout rules are conceptually appealing features of this framework. But a detailed model of this form needs to address the inconsistency of directly representing probability values with neural responses that are noisy (e.g., Sahani & Dayan, 2003). In addition, the responses of many visual neurons do not seem consistent with direct representation of posterior probability. For example, the shape of orientation tuning curves in area V1 neurons is preserved under changes in stimulus contrast. But lowering the stimulus contrast results in larger variance in the estimation of orientation, which implies a broadening of the posterior.

A widely followed alternative formulation represents probabilities implicitly, using the noisy responses of a population of neurons (e.g., Seung & Sompolinsky, 1993; Salinas & Abbott, 1994; Zhang et al., 1998; Zemel et al., 1998; Pouget et al., 2003; Jazayeri & Movshon, 2006; Ma, Beck, Latham, & Pouget, 2006).[6] Consider a population of N neurons whose responses represent the measurement in an optimal estimation problem. Suppose the mean firing rate of each neuron is determined by a tuning function $f_n(x)$, where x is the stimulus variable of interest. Suppose also that the number of spikes emitted by each of these neurons in a unit time interval to any given stimulus is statistically independent and follows a Poisson distribution. Then the joint likelihood function is the product of the individual Poisson probabilities:

$$P(\vec{r}\,|x) = \prod_n \frac{f_n(x)^{r_n}}{r_n!} e^{-f_n(x)}$$

where $\vec{r}$ is a vector containing the population spike counts.

At this point, one might use unsupervised (or supervised) regression to learn the optimal estimator from data without specifying or learning a prior probability. Alternatively, we can follow the Bayesian formulation, multiplying the likelihood by a prior $P(x)$, dividing by $P(\vec{r})$, and taking the (negative) log to write the log-posterior density:

$$-\log\left(P(x|\vec{r})\right) = \sum_n \log(r_n!) - \log P(\vec{r}) - \sum_n r_n \log\left(f_n(x)\right) + \sum_n f_n(x) - \log P(x)$$

Now consider the MAP estimator. The first two terms do not contain the stimulus variable, x, so we can drop them, arriving at an objective function that can be maximized over x to obtain the estimate:

$$E(x) = -\sum_n r_n \log\left(f_n(x)\right) + \sum_n f_n(x) - \log P(x)$$

The first term is a sum of the observed spike counts, weighted by the log tuning curve value of each neuron (Zhang et al., 1998; Jazayeri & Movshon, 2006). The second term is the sum of the tuning curves, and the third is the (negative) log of the prior. Much of the previous work on population coding has focused on the special case of orientation representation in V1 neurons or selectivity for motion direction in MT neurons, and in these cases the prior over orientation is typically assumed to be constant, as is the sum of the tuning curves (e.g., Zemel et al., 1998; Jazayeri & Movshon, 2006). These assumptions allow one to ignore the last two terms, and the resulting log-posterior objective function reduces to a simple weighted sum of spike counts, consistent with earlier proposals for linear readout (Bialek et al., 1991; Anderson & van Essen, 1994; Rieke et al., 1997). Note that later stages of processing (i.e., the estimator) presumably

would not have access to the tuning curves, and thus could not compute the linear weights directly by taking the log. Instead, the proper weights could again be learned using unsupervised (or supervised) regression. The linear representation of log probability is especially convenient for fusing independent sources of information, e.g., accumulating evidence over time (Jazayeri & Movshon, 2006; Beck et al., 2008) or combining evidence from multiple modalities (Ma et al., 2006). In these cases, one wishes to multiply the associated probability densities, which can be done by simply adding spike counts.

Rather than assuming a constant prior and sum of tuning curves, a more general solution could embed the prior into the measurements by arranging that the sum of the tuning curves is equal to the log of the prior:

$$\sum_n f_n(x) = \log P(x)$$

(Note: A precise form of this proposal would need to limit the smallest probability that could be represented.) Under these conditions, the last two terms cancel each other, and the full log-posterior may again be computed as a linear function (i.e., a weighted sum) of the spike counts. This is effectively a strategy for embedding the prior into the measurements, and is more efficient than encoding priors with spiking responses of another set of neurons (Ma et al., 2006). This solution would require that the brain adjust the tuning curves so as to sum to the log prior, which is essentially a resource allocation problem: Neurons should be adjusted so as to properly "cover" the distribution of inputs, assigning more resources (i.e., a higher total spiking response, which corresponds to an expenditure of more metabolic energy) to inputs that occur more frequently. This adjustment could be achieved by changing either their response gains, the overlap between their tuning curves, the widths of their tuning curves, or some combination of these.

In the probabilistic representation described above, a single optimal estimate can be computed by appropriately combining information over the entire population. Implementing this in the brain would presumably require creation of a redundant population of neurons to linearly recode the implicit representation into an explicit one (that is, a population whose responses equal the posterior or log posterior), from which a maximum (or mean) could be selected. Although this sort of explicit estimation has been assumed by many of the previously mentioned publications, it seems to me wasteful of neural resources, and not robust to the additional noise that would be introduced by neurons computing and representing the estimate. It seems more likely that the brain leaves the representation probabilistic and implicit, performing further calculations in a way that is consistent with this (e.g., Ma et al., 2006). Taking this *principle of delayed estimation* to its extreme, we could hypothesize that estimates need only be made explicit when the information reaches the motor system and the animal must execute an action (e.g., reaching out to grasp an object). At that point, the estimate is "computed" by a bone, which responds by moving according to the collective activity of all the muscle fibers that are pulling on it! One case that may be an exception to this is that of a binary estimate (i.e., a decision), for which the log-posterior need only be computed at two different values (rather than a continuum). Experimental evidence suggests that the firing rates of neurons in parietal cortex may represent such values directly (e.g., Platt & Glimcher, 1999; Gold & Shadlen, 2000).

Conclusions

Optimal estimation provides a formal framework for investigating and interpreting perceptual capabilities. The definition of the optimal estimator does not specify a fixed universal function but depends on four fundamental ingredients: the ensemble of input signals over which it is to be optimal, the (probabilistic) relationship between the signal and the measurement, the cost of misestimation, and the family of estimation functions from which the solution is to be chosen. In a biological system, these ingredients may change over time (especially the prior and loss function) and therefore presumably need to be learned and continually updated on the basis of recent input. I have reviewed three different formulations for developing an optimal estimator, each making different assumptions about the means by which these ingredients are obtained.

The basic formalism that I have presented here is highly oversimplified. In particular, I have sidestepped several important features of sensory systems that need to be incorporated in a full solution:

- Sensory computations occur in cascades of neural populations, and each of these presumably performs some transformations on the signals that are received from its afferents and introduces additional noise. The designation of a particular population as the "measurement" is therefore somewhat artificial.
- Sensory computations occur over time. Many optimal estimation problems can be rewritten in a form that can be computed incrementally (the classical solution in statistical signal processing is known as the *Kalman filter*), and such solutions have been used to model temporal aspects of neural processing (e.g., Rao & Ballard, 1997; Denève, Duhamel, & Pouget, 2007).
- Many sensory inference computations depend on information beyond the measurements and the prior, such as inputs from multiple sensory modalities or feedback in the form of attentional signals, or from emotional or cognitive centers. It might be possible to formalize these effects as a contextual form of prior (Jaynes, 2003).

In summary, the challenge for future research is to develop optimal estimation solutions that can be plausibly mapped onto brain architecture, that are flexible and adaptive, that may be cascaded (with additional noise introduced at each stage of the cascade), and that can be learned in a primarily unsupervised setting. The time seems ripe for this. A long tradition of rigorous study in statistical inference has been developed and refined into engineering tools. The use and development of these have recently been accelerated by new methods and algorithms that have arisen in the machine learning community. Coupling these with a new generation of experimental measurement technologies (especially those for obtaining responses from groups or populations of neurons simultaneously) leads me to conclude that we are on the verge of fundamental advances in our understanding of sensory processing.

ACKNOWLEDGMENTS Thanks to Alan Stocker, Mehrdad Jazayeri, Martin Raphan, and the section editors, Tony Movshon and Brian Wandell, for helpful comments and suggestions. This work was financially supported by the Howard Hughes Medical Institute, the National Institutes of Health (EY018003), and the Sloan-Swartz Center for Theoretical Visual Neuroscience at New York University, but the views expressed herein are my own.

NOTES

1. Throughout, x and/or m can be scalar-valued or vector-valued quantities. For example, m might represent responses of a population of neurons.
2. A more sophisticated solution would adjust the smoothness of the estimator adaptively to the amount of data. This issue of model complexity and its relationship to learning from data is fundamental to the study of machine learning.
3. Bias can arise because the estimator is restricted to lie in a family that does not include the best solution. But it can also arise due to asymmetries in the posterior (e.g., due to the influence of a prior) or cost function, in which case it should be viewed as desirable. E. T. Jaynes (2003) described the insistence on unbiased estimators as a "pathology of orthodox statistics."
4. By presenting the Bayesian and regression formulations side by side, I hope to alleviate some of the tensions between the Bayesian and "frequentist" viewpoints. (See Jaynes, 2003, for further discussion.)
5. A less well-known form of empirical Bayes estimation arises from rewriting the estimator directly in terms of the distribution of measurements, which can be approximated from the observed data (Miyasawa, 1956). This form may be derived from the prior-free estimator described previously (Raphan & Simoncelli, 2007).
6. Note that this formulation is encoding the uncertainty due to the noise in the population response, but not the noise or structural ambiguities in the input (e.g., Sahani & Dayan, 2003).

REFERENCES

Anderson, C., & van Essen, D. (1994). Neurobiological computational systems. In *IEEE World Congress on Computational Intelligence*. New York: IEEE Press.

Barber, M. J., Clark, J., & Anderson, C. H. (2003). Neural representation of probabilistic information. *Neural Comput.*, *15*(8), 1843–1864.

Barlow, H. B. (1980). The absolute efficiency of perceptual decisions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *290*, 71–82.

Beck, J., Ma, W. J., Kiani, R., Hanks, T., Churchland, A., Roitman, J., Shadlen, M., Latham, P., & Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, *60*(6), 1142–1152.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., & Warland, D. (1991). Reading a neural code. *Science*, *252*, 1854–1857.

Bialek, W., & van Steveninck, R. D. (2005). *Features and dimensions: Motion estimation in fly vision.* Technical Report qbio/0505003. Available online at: http://arxiv.org/.

Denève, S., Duhamel, J.-R., & Pouget, A. (2007). Optimal sensorimotor integration in recurrent cortical networks: A neural implementation of Kalman filters. *J. Neurosci.*, *27*(21), 5744–5756.

Eliasmith, C., & Anderson, C. H. (2002). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.*, *9*, 292–303.

Foldiak, P. (1993). The "ideal homunculus": Statistical inference from neural population responses. In F. H. Eeckmann & J. M. Bower (Eds.), *Computation and neural systems* (pp. 55–60). Norwell, MA: Kluwer Academic.

Geisler, W. S. (1989). Ideal-observer theory in psychophysics and physiology. *Phys. Scripta*, *39*, 153–160.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Nature*, *233*, 1416–1419.

Glimcher, P. W., Camerer, C., Poldrack, R., & Fehr, E. (Eds.). (2008). *Neuroeconomics: Decision making and the brain.* London: Academic Press.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*, 390–394.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley.

Hinton, G., & Sejnowski, T. J. (Eds.). (1999). *Unsupervised learning: Foundations of neural computation*. Cambridge, MA: MIT Press.

Hinton, G. E. (1992). How neural networks learn from experience. *Sci. Am.*, *267*(3), 144–151.

Jaynes, E. T. (1957). *How does the brain do plausible reasoning?* Technical Report 421. Stanford University, Microwave Laboratory Technical Report. Reprinted in G. J. Erickson and C. R. Smith (Eds.) (1988). Maximum-entropy and Bayesian methods in science and engineering (Vol. 1, pp. 1–24). Dordrecht: Kluwer.

Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge, UK: Cambridge University Press.

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nat. Neurosci.*, *9*(5), 690–696.

Kersten, D. (1990). Statistical limits to image understanding. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 32–44). Cambridge, UK: Cambridge University Press.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol.*, *55*, 271–304.

Knill, D., & Richards, W. (1996). *Perception as Bayesian inference.* Cambridge, UK: Cambridge University Press.

Knill, D. C., Field, D., & Kersten, D. (1990). Human discrimination of fractal images. *J. Opt. Soc. Am. [A]*, *7*, 1113–1123.

Körding, K. P. (2007). Decision theory: What "should" the nervous system do? *Science*, *318*(5850), 606–610.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci.*, *9*, 1432–1438.

Maloney, L. T. (2002). Statistical decision theory and biological vision. In D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world: Psychological and philosophical issues in perception* (pp. 145–189). New York: John Wiley.

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of visual perception: Testing Bayesian transfer. *Vis. Neurosci.* (special issue). In press.

Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modeling of visual perception. In R. Rao, M. Lewicki, & B. Olshausen (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 13–36). Cambridge, MA: MIT Press.

Miyasawa, K. (1956). An empirical Bayes estimator of the mean of a normal population. *Bull. Inst. Int. Statist.*, *38*, 181–188.

Paninski, L. (2006). Nonparametric inference of prior probabilities from Bayes-optimal behavior. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (Vol. 18, pp. 1067–1074). Cambridge, MA: MIT Press.

Pelli, D. G., & Farell, B. (1999). Why use noise? *J. Opt. Soc. Am. [A]*, *16*, 647–653.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*, 233–238.

Potters, M., & Bialek, W. (1994). Statistical mechanics and visual signal processing. *J. Phys. France*, *4*, 1755–1775.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annu. Rev. Neurosci.*, *26*, 381–410.

Rao, R. P. N., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Comput.*, *9*, 721–763.

Raphan, M., & Simoncelli, E. P. (2007). Learning to be Bayesian without supervision. In B. Schölkopf, J. Platt, & T. Hofmann (Eds.), *Advances in neural information processing systems* (Vol. 1, pp. 1145–1152). Cambridge, MA: MIT Press.

Rieke, F., & Baylor, D. (1998). Single photon detection by rod cells of the retina. *Rev. Modern Phys.*, *70*, 1027–1036.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.

Sahani, M., & Dayan, P. (2003). Doubly distributional population codes: Simultaneous representation of uncertainty and multiplicity. *Neural Comput.*, *15*, 2255–2279.

Salinas, E., & Abbott, L. F. (1994). Vector reconstruction from firing rates. *J. Comp. Neurosci.*, *1*, 89–107.

Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *J. Neurophysiol.*, *76*(4), 2790–2793.

Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neural population codes. *Proc. Natl. Acad. Sci. USA*, *90*, 10749–10753.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.*, *16*(4), 1486–1510.

Simoncelli, E. P. (1993). *Distributed analysis and representation of visual motion*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA. (Also available as MIT Media Laboratory Vision and Modeling Technical Report #209.)

Simoncelli, E. P. (2003). Local analysis of visual motion. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1616–1623). Cambridge, MA: MIT Press.

Snippe, H. P. (1996). Parameter extraction from population codes: A critical assessment. *Neural Comput.*, *8*(3), 511–530.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, *9*(6), 1135–1151.

Stocker, A. A., & Simoncelli, E. P. (2006a). Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.*, *9*(4), 578–585.

Stocker, A. A., & Simoncelli, E. P. (2006b). Sensory adaptation within a Bayesian framework for perception. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems (NIPS*05)* (Vol. 18, pp. 1291–1298). Cambridge, MA: MIT Press.

von Helmholtz, H. (1925). *Treatise on physiological optics* (Vol. 3). New York: Optical Society of America.

Weiss, Y., & Fleet, D. J. (2002). Velocity likelihoods in biological and machine vision. In R. Rao, B. Olshausen, & M. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 81–100). Cambridge, MA: MIT Press.

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Comput.*, *10*, 403–430.

Zhang, K., Ginzburg, I., McNaughton, B., & Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *J. Neurophysiol.*, *79*, 1017–1044.

V

MOTOR SYSTEMS

Chapter 37 BIZZI AND MUSSA-IVALDI 541

38 DUM AND STRICK 553

39 GRAYBIEL AND MINK 565

40 SHADMEHR AND KRAKAUER 587

41 MULLIKEN AND ANDERSEN 599

42 TODOROV 613

43 RIZZOLATTI, FOGASSI, AND GALLESE 625

44 GRAFTON, AZIZ-ZADEH, AND IVRY 641

Introduction

SCOTT T. GRAFTON AND EMILIO BIZZI

HUMANS MOVE with a purpose. Thus the study of motor systems requires more than just the description of movement control. It demands a broad perspective based on anatomical, functional, and computational principles to describe how the nervous system can enable goal-directed behavior. The challenge in solving this problem remains daunting because the intrinsic neural codes that are used to generate motor actions are not fully known, the principles that lead to optimal performance are only just beginning to be modeled, and the system is nonstationary in that adaptation and skill learning appear continuously and on many time scales. The section on motor systems addresses this challenge by the inclusion of complementary approaches, spanning many levels of analysis.

The section begins with a review of motor primitives by Bizzi and Mussa-Ivaldi. This is a critical solution to the degrees-of-freedom problem in limb movement. There are an infinite number of possible muscle activation patterns that could lead to similar movements. How does the nervous system constrain this redundancy? Motor primitives simplify the problem by reducing muscle activations to a set of muscle groups that can be combined, analogous to the mixing of basis functions, to generate characteristic limb movements.

Anatomical studies in the chapter by Dum and Strick extend their previous work demonstrating distinct output channels from different parts of the basal ganglia to distinct cortical areas. New work reveals distinct output channels from the cerebellar nuclei to cortex as well. The existence of distinct cortical projects from basal ganglia and the cerebellum has many important functional implications and could explain the diverse semiology of deficits in patients with lesions to either the basal ganglia or the cerebellum. Alternative models of basal ganglia function, based on both clinical and physiological evidence, are reviewed in greater

detail in the chapter by Graybiel and Mink. They present recent evidence for dopamine mediated learning mechanisms, action selection processes and on-line control within the basal ganglia. Particular emphasis is placed on the selection of action based on prior experience combined with contextual information that might influence choice.

Shadmehr and Krakauer consider the problem of action dynamics more broadly. They consider patients with lesions of the cerebellum, parietal cortex, and basal ganglia and interpret their deficits in terms of computational processes such as state estimation, optimization, prediction, cost, and reward. From this evidence, they argue for relative specialization by which the cerebellum builds internal models that predict sensory outcome of motor commands and correct motor commands, the parietal cortex is used for state estimation, and the basal ganglia are needed for learning costs and rewards associated with sensory states.

Evidence for a unique role of the parietal cortex in state estimation, that is, the integration of vision and somatosensory information with motor command is considered in the chapter by Mulliken and Andersen. They emphasize the role of the parietal cortex for generating a forward model that predicts the sensory consequences of a movement. This prediction is likely to be used in a number of action-relevant processes, including on-line control, evaluating performance with a desired outcome, canceling reafferent input, and mental simulation and determining agency.

The computational methods used in state estimation merge sensory feedback and ongoing motor commands into a common theoretical framework. These different sources of information are traditionally considered to represent solutions to very different computational problems. Perception has to do with inferring the state of the world given sensory data and action with generating motor commands that lead to a task goal. In his chapter, Todorov summarizes how these two problems are in many ways related. Sensory inference and motor prediction can be united in a single computational framework. With this, it is possible to explore computational similarities and differences between the two systems.

The overlap between perception and action at the neural level is considered in the chapter by Rizzolatti, Fogassi, and Gallese. Using the mirror neuron as a core mechanism for representing an action that is either executed or perceived, they present evidence that extends this general mechanism to humans, where it could be used to understand intentions in others. Impairment of this process might explain some of the clinical deficits that are seen in autism spectrum disorders.

In the final chapter in this section, Grafton, Aziz-Zadeh, and Ivry consider the importance of hierarchical representation as an organizing principle for understanding how people are capable of creating as well as recognizing the meaning of complex, goal-oriented action. They review functional and behavioral evidence in human experiments of hand-object interactions, bimanual control, and the integration of semantics into action planning. The results support the existence of a highly flexible control hierarchy rather than a strict anatomical hierarchy for organizing complex motor behavior.

37 Neurobiology of Coordinate Transformations

EMILIO BIZZI AND FERDINANDO A. MUSSA-IVALDI

ABSTRACT A broad variety of motor plans and concomitant control actions can be expressed by the superposition of force fields representing the mechanical effects of motor synergies. This principle of superposition may lead to the execution of motor plans by controlling the nonlinear dynamics of the body in the presence of redundant muscles and degrees of freedom.

From movement planning to execution

A critical issue in the generation of motor behavior concerns the hierarchical organization of movement planning and movement execution. This concept is derived from engineering notions of modular control by which the problem of movement is decomposed into subproblems that can be addressed separately. One of the great complications of the movement control problem is that any given goal can be reached by a multiplicity of means. If the goal is to move the hand from point A to point B, a variety of paths can be chosen; a variety of trajectories in joint space can be utilized to realize the path. As an example, consider the simple task of reaching for a glass of water on a table. To reach for it, the brain must generate a temporal sequence of activations of the arm muscles. The pattern of neural impulses that controls the contraction of each muscle can be thought of as "coordinates" in an abstract geometrical space (Holdefer & Miller, 2002). In this space, the goal of reaching the glass can be represented as a point whose coordinates are the muscle activations needed to perform the appropriate reach. What happens to these motor coordinates—and to the goal of reaching the glass in the space of muscle activities—if our body moves to another position, such as from standing to sitting? To reach the glass, the arm must now move in a different way, and the muscles must be driven by different commands. As a consequence, the coordinates of the goal of reaching the glass in the space of muscle activities are changed. This is what mathematicians call a *coordinate transformation*, a computation that can be of quite considerable complexity, given the many different ways in which one may have to activate the muscles to reach the same point in space.

A basic concept in many fields of science, including systems-level neuroscience, is the concept of a coordinate system (Bishop & Goldberg, 1980). A coordinate system is a system of numbers that, taken together, identify the location of a point in space. The space could be the ordinary three-dimensional space in which we move, or it could be an abstract space with a larger or even infinite number of dimensions that may be placed in correspondence with a physical system. For example, the posture of a marionette may be represented by specifying each of its joint angles on a separate axis. Thus, the joint angles provide a coordinate system for the marionette. The state of a biological system, such as the human arm, is described within the nervous system by the collection of neural activities that constitute incoming sensory signals and outgoing motor commands. Although there are several possible coordinate systems—actually an infinite number—to describe different sensory and motor signals, these coordinate systems fall quite naturally into three classes: neuromuscular coordinates, joint coordinates, and endpoint coordinates.

ENDPOINT COORDINATES Endpoint coordinates are appropriate for describing the goal of an action and the interaction with the environment. These coordinates may capture the highly regular properties of reaching behavior when the location of the hand is rendered in Cartesian space (Morasso, 1981; Soechting & Lacquaniti, 1981). Morasso (1981) instructed human subjects to point with one hand to different visual targets that were randomly activated (figure 37.1). His analysis of the movements showed two kinematic invariances: (1) The hand trajectories were approximately straight segments, and (2) the speed profile or tangential velocity of the hand for different movements always appeared to have a bell-shaped configuration, as the time needed to accelerate the hand was approximately equal to the time needed to bring it back to rest. Because these simple and invariant features were detected at a different shoulder and elbow angles, these results suggest that planning by the central nervous system (CNS) takes place in terms of hand

EMILIO BIZZI McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts
FERDINANDO A. MUSSA-IVALDI Department of Physiology, Northwestern University Medical School, Chicago, Illinois

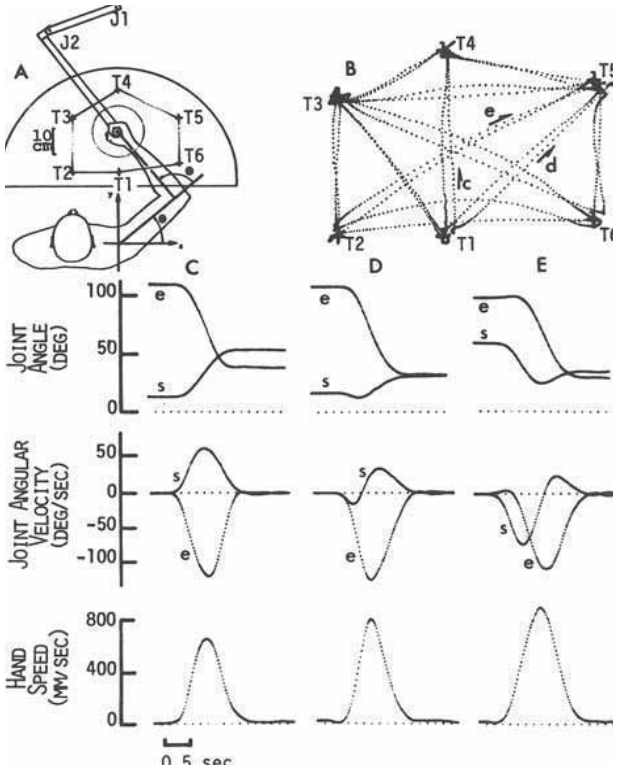


FIGURE 37.1 (*A*) Plan view of a seated subject grasping the handle of the two-joint hand-position transducer (designed by N. Hogan). The right arm was elevated to shoulder level and moved in a horizontal work space. Movement of the handle was measured with potentiometers located at the two mechanical joints of the apparatus (J1, J2). The subject was positioned so that J1 lay on the *Y*-axis. A horizontal semicircular plate located just above the handle carried the visual targets. Six visual target locations (T1–T6) are illustrated as crosses. The digitized paths between targets and the curved path were obtained by moving the handle along a straight edge from one target to the next and then along a circular path; movement paths were reliably reproduced. (*B*) A series of digitized handle paths (sampling rate: 100 Hz) performed by one subject in different parts of the movement space. The subject moved his hand to the illuminated target and then waited for the appearance of a new target. Targets were presented in random order. *Arrows* show direction of some of the hand movements. (*C*, *D*, *E*) Kinematic data for three of the movements the paths of which are shown in *B*. Letters show correspondence (e.g., data under *C* are for path c in *B*). Abbreviations: e, elbow joint; s, shoulder (angles measured as indicated in *A*). (From Bizzi E., and Abend, W. K., 1983, *Motor Control Mechanisms in Health and Disease*, J. E. Desmedt, ed. New York: Raven Press.)

motion in space. Here, the hand is regarded almost as a disembodied object whose movements are planned by the CNS independently of the geometrical and mechanical properties of the arm and its muscles. Morasso's observations were extended to more complex curved movements performed by human subjects in an obstacle avoidance task (Abend, Bizzi, & Morasso, 1982). Again, kinematic invariances were present in the hand and not in the joint motion. Later, Flash and Hogan (1985) showed that the kinematic behavior described by Morasso (1981) and Abend and colleagues (1982) could be derived from a single organizing principle based on optimizing endpoint smoothness. A different view was offered by Kawato and coworkers (Uno, Kawato, & Suzuki, 1989), who proposed that the shape of movement trajectories is not explicitly planned but is rather a side effect of the CNS optimizing a dynamical variable: the rate of change of joint torque. However, the observation that movement trajectories are recovered when the movement is perturbed by a force field (Shadmehr & Mussa-Ivaldi, 1994) reinforced the idea that the CNS recovers the trajectory of the hand by producing the torques that are needed to compensate for the external forces. This supports the notion of planning in endpoint coordinates.

MUSCLE COORDINATES Muscle coordinates (Holdefer & Miller, 2002) afford the most direct representation for the motor output of the central nervous system. A position in this coordinate system may be expressed, for example, as a collection of muscle lengths. Accordingly, a force vector in the same coordinate system is a collection of muscle tensions. The number of actuator coordinates depends upon how detailed is the model of control under consideration. Unlike generalized coordinates, actuator coordinates do not constitute a system of mechanically independent variables; one cannot set arbitrary values to muscle lengths without eventually violating a kinematic constraint.

At the most detailed level, individual motor units may be considered as actuator elements. In this case, the order of magnitude of the actuator space can be of the order of 10 or 100,000 dimensions.

JOINT COORDINATES AND THE GENERATION OF FORCES A different way of describing body motions is to provide the set of joint angles that define the orientation of each skeletal segment either with respect to fixed axes in space or with respect to the neighboring segments. This is a well-known representation in robotics and analytical mechanics called *configuration space* (Spong, Hutchinson, & Vidyasagar, 2005). Joint angles are a particular instance of generalized coordinates. Generalized coordinates are independent variables that are suitable for describing the dynamics of a system (Goldstein, 1980; Jose & Saletan, 1998). In particular, the dynamics of limbs such as the human arm are described by systems of coupled differential equations relating the generalized coordinates to their first and second time derivatives and to the generalized forces.

In vector notation, the dynamics equations for a multi-jointed limb are succinctly written as

$$I(q)\ddot{q} + G(q, \dot{q}) + E(q, \dot{q}, t) = C(q, \dot{q}, u(t)) \qquad (1)$$

where $q = (q_1, q_2, \ldots, q_N)$ is the limb configuration in joint-angle coordinates, $\dot{q}$ and $\ddot{q}$ are respectively the first (velocity) and second (acceleration) time derivatives of q, I is a $N \times N$

matrix of inertia (that is configuration-dependent), $G(q, \dot{q})$ is a vector of centripetal and Coriolis torques (Spong et al., 2005), and $E(q, \dot{q}, t)$ is a vector of external torques, which, in general, depends upon the state of motion of the limb and upon time. The whole left-hand side of equation 1 represents the torque due to the inertial properties of the arm and to the action of the environment (part of which may be considered "noise"). The term $C(\cdot)$ on the right-hand side stands for the net torque generated nonlinearly by the muscles, by the environment (e.g., the gravitational torque), and by other dissipative elements, such as friction. The time-function $u(t)$ is a control vector—for example, a set of neural signals directed to the motoneurons or a representation of a desired limb position at time t. The left-hand side of equation 1 represents the passive dynamics associated with limb inertia. The right-hand side is the applied force, which is the output of a control process. An additional term, not represented in equation (1), is the noise associated with the control signal.

The way in which the CNS implements the dynamic equation (1) has been the focus of a number of investigations in the last 20 years. Through work in humans and monkeys, a number of investigators have put forward the hypothesis that the CNS generates movements and forces by using internal models of the limbs and of the environment with which they come into contact (Thoroughman & Taylor, 2005; Sabes, 2000; Davidson & Wolpert, 2004, Flanagan & Wing, 1997; Kawato & Wolpert, 1998; Krakauer, Ghilardi, & Ghez, 1999; Shadmehr & Mussa-Ivaldi, 1994; Wolpert, Miall, & Kawato, 1998).

Alternative proposals have been made that do not depend on the solution of complicated inverse dynamics problems (Feldman & Latash, 2005). Specifically, it has been proposed that the CNS may transform the desired hand motion into a series of equilibrium positions (Bizzi, Accornero, Chapple, & Hogan, 1984). The forces that are needed to track the equilibrium trajectory result from the intrinsic elastic properties of the muscles and from local feedback loops (Feldman, 1966; Bizzi, Polit, & Morasso, 1976; Hogan, 1984). However, if movements are to be executed with any substantial amount of acceleration, then the inertial forces of the arm have to be taken implicitly into account, as they tend to displace the limb away from the equilibrium trajectory. In this case, some computational mechanism is required to derive the equilibrium trajectory that is adequate to move the hand along the desired trajectory. It is then merely a semantic issue to distinguish this operation from an explicit dynamics computation.

Another view on solving the dynamic equations and to execute a motor plan is the hypothesis that motor behavior of vertebrates is based on simple units (motor primitives) that can be flexibly combined to accomplish a variety of motor tasks. Here is an apt and succinct quote from Cvitanovic (2000), concerning the general problems posed by complex nonlinear dynamics:

> Armed with a computer and a great deal of skill, one can obtain a numerical solution to a nonlinear partial differential equation. The real question is; once a solution is found, what is to be done with it? . . . Dynamics drives a given spatially extended system through a repertoire of unstable patterns; as we watch a "turbulent" system evolve, every so often we catch a glimpse of a familiar pattern. For any finite spatial resolution, the system follows approximately for a finite time a pattern belonging to a finite alphabet of admissible patterns, and the long-term dynamics can be thought of as a walk through the space of such patterns, just as chaotic dynamics with a low-dimensional attractor can be thought of as a succession of nearly periodic (but unstable) motions.

The key concept here is that complex behavior may be analyzed by a combination of patterns.

Building blocks for computation of dynamics: Compositionality

In the natural world some complex systems are discrete combinatorial systems—they utilize a finite number of discrete elements to create larger structures. The genetic code and language phenomena are examples of systems in which discrete elements and a set of rules can generate a large number of meaningful entities that are quite distinct from those of their elements. A question of considerable importance is whether this fundamental characteristic of language and genetics is also a feature of other biological systems. In particular, the question is whether the activity of the vertebrate motor system, with its impressive capacity to find original motor solutions to an infinite set of ever-changing circumstances, results from the combinations of discrete elements.

The ease with which we move hides the complexity inherent in the execution of even the simplest tasks. Even movements we make effortlessly, such as reaching for an object, involve the activation of many thousands of motor units in numerous muscles. Given this large number of degrees of freedom of the motor system, a number of investigators have put forward the hypothesis that the CNS handles this large space with a hierarchical architecture based upon the utilization of discrete building blocks whose combinations result in the construction of a variety of different movements (Arbib, 1981; Tsetlin, 1973). In particular, investigators influenced by the AI perspective on the control of complex systems have argued for a hierarchical decomposition with modules, or building blocks, as the most effective way to select a control signal from a large search space (Russell & Norvig, 1995).

The construction of natural motor behavior with muscle synergies

For a long time, investigators have recognized that one of the basic questions in motor performance is whether the cortical motor areas control individual muscles or make use

of synergistically linked groups of muscles (Macpherson, 1991). Given that no natural movement involves just one muscle, any motor act, a fortiori, involves a "muscle synergy," the question, then, has been whether the synergistic activation of muscles derives from a fixed common neural drive or is merely a phenomenological event of a given motor coordination. The recent introduction of novel computational procedures to extract synergies from large sets of EMG signals in intact behaving animals and humans has opened a different way to approach the issue of synergies (Soechting & Lacquaniti, 1989; Tresch, Saltiel, d'Avella, & Bizzi, 2002; Saltiel, Wyler-Duda, d'Avella, Tresch, & Bizzi, 2001; d'Avella & Tresch, 2002; d'Avella, Portone, Fernandez, & Lacquaniti, 2006).

In 1999, Tresch, Saltiel, and Bizzi developed a computational method to extract synergies from the recorded muscle activations. This method is based on decomposing the observed muscle patterns as simultaneous combinations of a number of synergies. This decomposition is obtained by using an iterative algorithm that is initialized with a set of arbitrary synergies. The nonnegative weighting coefficients of these arbitrary synergies that best predict each response are then found. The synergies are then updated by minimizing the error between the observed response and the predicted response. This process is iterated until the algorithm converges on a particular set of synergies. The algorithm extracts both a set of synergies and the weighting coefficients of each synergy used to reconstruct the electromyelogram responses.

Note that there are a number of factorization algorithms to assess the hypothesis that motor behavior might be produced through a combination of a small number of synergies. Tresch, Cheung, and d'Avella (2006) have compared different algorithms and found that in general, most of the algorithms that are used to identify muscle synergies, such as nonnegative matrix factorization, independent component analysis, and factor analysis, perform comparably. When they applied these methods to experimentally obtained data sets, the best-performing algorithms identified synergies that were very similar to one another. These results suggest that the muscle synergies found by a particular algorithm are not an artifact of that algorithm but reflect basic aspects of muscle activation.

The coordination among muscle recruitments expressed by a synergy might also be extended to the temporal domain. This idea has led to the introduction of a novel factorization algorithm (d'Avella & Tresch, 2002) to extract time-varying synergies, that is, the coordinated activations of groups of muscles with specific time-varying profiles. Time-varying synergies can naturally capture specific asynchronous activations of groups of muscles and provide a parsimonious model for the generation of muscle patterns. In fact, once the synergies have been specified, one amplitude scaling coefficient and one time delay coefficient per synergy are sufficient for generating a muscle pattern. In contrast, the entire time course of the weighting coefficients is required with synchronous synergies.

To directly assess this hypothesis, d'Avella, Saltiel, and Bizzi (2003) examined several motor behaviors in intact, freely moving frogs by recording simultaneously from a large number of hindlimb muscles during locomotion, swimming, jumping, and defensive reflexes (d'Avella et al., 2003; d'Avella & Bizzi, 2005). An iterative algorithm was used to decompose the muscle patterns as combinations of time-varying muscle synergies independently scaled in amplitude and shifted in time. This iterative algorithm finds a set of muscle synergies and, for each muscle pattern, the amplitude and delay of each synergy that minimize the reconstruction error for the entire data set. Figure 37.2 shows the five time-varying synergies that were extracted from all the rectified, low-pass filtered, and integrated (10 ms) EMGs recorded during a total of 2174 jumps, walking cycles, and swimming cycles in three frogs. The five extracted synergies include all 13 muscles. The first three synergies (W_1, W_2, and W_3) recruit mainly extensors, while W_4 and W_5 recruit mainly flexors. The most active muscles of synergy W_1 are the hip extensors rectus internus (RI), adductor magnus (AD), and semimembranosus (SM); the knee extensor vastus internus (VI); and the ankle extensors peroneus (PE) and gastrocnemius (GA). The most active muscles of the synergy W_2 are SM, vastus externus (VE), and GA. In W_3, RI, SM, and VI are the most active. The flexors dominate synergy W_4 with rectus anterior (RA), biceps (BI), and iliopsoas (IP). Synergy W_5 includes mainly semitendinosus (ST) and IP. Note that some of the muscles are present in more than one synergy. The R^2 for the five synergies extracted from the entire data set was 0.78. Thus a large fraction of the total variation of the data was described by a model that has just 10 parameters (five amplitude and five timing coefficients) once the synergies have been determined.

Figure 37.3 shows the reconstruction of muscle patterns (rectified, filtered, and integrated EMGs, *thin line and shaded area*) for a jump, a cycle of walking, and a cycle of swimming as combinations of the five synergies shown in figure 37.2 (*thick line*). The synergy's amplitude is shown as the height of the rectangles below the EMGs, and the delay coefficients are shown by their horizontal position. The essential features of the three muscle patterns are well captured by scaling in amplitude and shifting in time the five time-varying synergies. In jumping (first column), two of the three extension synergies are active (W_1 and W_3) together with the two flexion synergies (W_4 and W_5). In walking (second column), synergies 1, 3, and 4 appear again, but their amplitude balance and recruitment order are radically different from jumping: W_4 dominates in amplitude, while W_1 and W_3 are

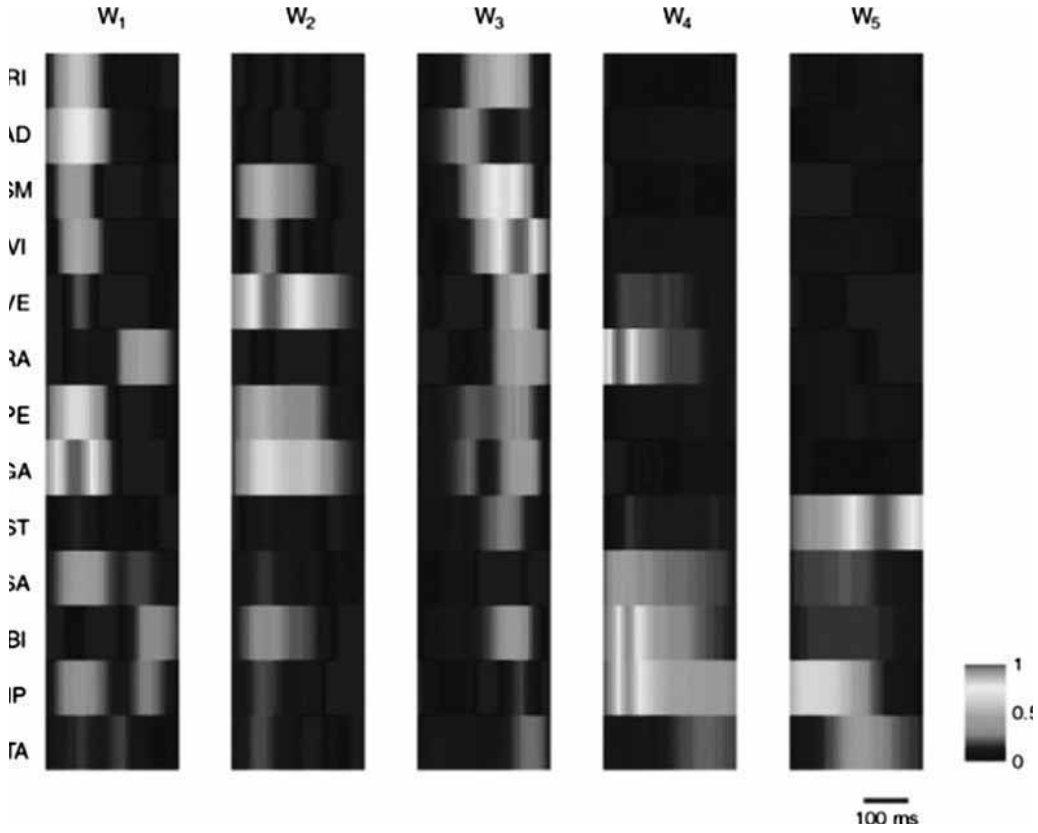


FIGURE 37.2 Time-varying muscle synergies extracted from jumping, swimming, and walking muscle patterns in three frogs. Each synergy (columns W_1 to W_5) represents the activation time course (in color code) of 13 muscles over 30 samples (total duration: 300 ms) normalized to the maximum sample of each muscle. Abbreviations: RI, rectus internus; AD, adductor magnus; SM, semimembranosus; VI, the knee extensor vastus internus; VE, vastus externus; RA, rectus anterior; PE, the ankle extensors peroneus; GA, gastrocnemius; ST, mainly semitendinosus; SA, semitendinosus; BI, biceps; IP, iliopsoas; TA, tibialis anterior. (From Bizzi, E., Cheung, V. C., d'Avella, A., Saltiel, P., Tresch, M. C., 2008, Combining modules for movement, *Brain Res. Rev.*, *57*, 125–133.) (See color plate 51.)

relatively small, and the timing between W_3 and W_4 is reversed. Swimming (third column), in contrast, is dominated by W_2 and to a lesser extent by W_5.

The examples illustrated by figure 37.3 demonstrate two important points: (1) that the same synergies are found in different behaviors and (2) that different behaviors may be constructed by combining the same synergies with different timing and amplitude.

The examples illustrated in figures 37.2 and 37.3 address the important question of whether the synergies extracted by a computational procedure have biological standing. A compelling criterion is the presence of the same synergy, with its own internal temporal structure, in different behaviors, as illustrated in figure 37.3. Additional evidence supporting the idea that synergies are indeed functional units was shown by Giszter and Kargo (2000). They showed examples of deletions as well as of additions during leg motions in the frog. Kargo and Giszter (2000) showed that the spinalized frog is able to produce corrective movements in response to unexpected perturbations. They placed an obstacle in the path of the leg and showed that when the leg hit the obstacle, the added synergy was one of the sets of six previously identified synergies.

Other investigators have generated corroborative evidence for modular organization in cats (Lemay, Galagan, Hogan, & Bizzi, 2001; Ting & Macpherson, 2005; Krouchev, Kalaska, & Drew, 2006; Torres-Oviedo et al., 2006) and the turtle (Stein, Oguztoreli, & Capaday, 1986; Stein, McCullough, & Currie, 1998). In addition, results from the study of the muscle patterns during reaching in humans (d'Avella et al., 2006) suggest that this is a general strategy used by all vertebrates for simplifying the control of limb movements.

A clear-cut example of a recombination of synergies is from locomotion with the different limb central pattern generators (CPGs). Each CPG can operate independently, but the four limb CPGs can also be combined in different patterns as in a walk, a trot, or a gallop. On the basis of extensive indirect evidence, Grillner (1981, 1985) suggested that

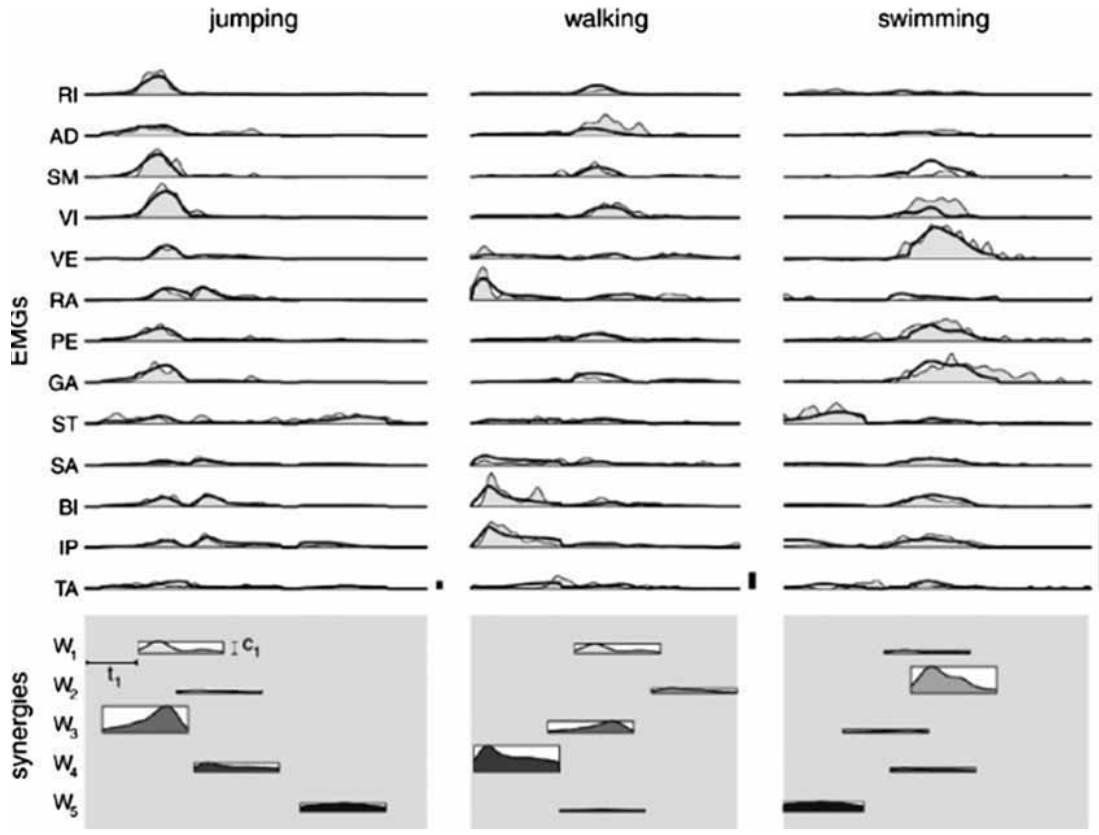


FIGURE 37.3 Examples of reconstruction of EMG patterns as combinations of time-varying muscle synergies. The three columns are examples of a jump, a walking cycle, and a swimming cycle. *Upper section*: The thick line shows the reconstruction of muscle patterns and the shaded area represents the rectified, filtered and integrated EMGs. *Lower section*: The coefficients of the five synergies as the horizontal position (onset delay, t_i) and the height (amplitude, c_i) of a rectangle whose width corresponds to the synergy duration. The shaded profile in each rectangle illustrates the averaged time course of the muscle activation waveforms of the corresponding synergy. Note the different amplitude scaling used in the three columns. (From Bizzi, E., Cheung, V. C., d'Avella, A., Saltiel, P., Tresch, M. C., 2008, Combining modules for movement, *Brain Res. Rev.*, 57, 125–133.)

each limb CPG can be further subdivided into unit CPGs that control synergist muscles acting at each joint. It has also been proposed that these different unit CPGs or synergies can be the independent target for the supraspinal commands used to design different volitional movements involving a limited set of joints (Grillner, 1985, 2006; Grillner & Zangger, 1979).

In conclusion, the evidence provided by studies from different laboratories and in different species indicates that combining muscle synergies is a strategy that the CNS utilizes for the construction of movements in vertebrates.

Physiological basis of muscle synergies: Modularity in the frog spinal motor system

With microstimulation of the spinal interneuronal regions, Bizzi, Mussa-Ivaldi, and Giszter (1991), Giszter, Mussa-Ivaldi, and Bizzi (1993), and Tresch and colleagues (1999) have provided evidence for a modular organization of the frog's and rat's spinal cord. These experiments found that only a few distinct types of motor outputs could be evoked by either electrical (Bizzi et al., 1991) or NMDA stimulation (Saltiel et al., 2001). Importantly, when stimulation was applied simultaneously to two different sites in the spinal cord, each of which when stimulated produced a different motor output, the resulting motor output was a simple linear combination of the separate motor outputs (Mussa-Ivaldi, Giszter, & Bizzi, 1994; Lemay et al., 2001). In subsequent experiments, Tresch and colleagues (1999) showed that the motor response evoked from cutaneous stimulation of a particular site on the hindlimb resulted from the weighted combination of a few muscle synergies. When Tresch and colleagues (1999) compared the distinct muscle synergies derived from cutaneous stimulation with the patterns of muscle activation evoked by microstimulation of the frog spinal cord, he found that the two sets of EMG responses were very similar to one another. In addition, the synergies evoked by NMDA were found by Saltiel and colleagues (2001) to be qualitatively similar to those described by Tresch and colleagues (1999). Taken together, these experiments

have provided evidence in support of a spinal modular organization underlying natural behaviors produced by the frog spinal cord.

Translating plans into actions: The mechanical basis for compositionality

The microstimulation studies described above have revealed that the mechanical effect of a muscle synergy is captured by a field of forces that vary both with space and in time. The combination of force fields generated by multiple synergies offers the CNS a way for solving the problem of dynamics (equation 1) in the presence of motor redundancy. Motor redundancy is a common property of multijoint limbs, such as the human arm, in which there are more joint angles than endpoint variables and more muscles than joint angles. Many of the computational problems associated with redundancy are removed by expressing a motor plan as a force field in endpoint coordinates and by approximating this field with a superposition of force fields corresponding to muscle synergies. This is illustrated in more detail by the following argument.

The graph in figure 37.4*A* highlights the main challenge associated with redundancy of the musculoskeletal apparatus: The transformations between actuator, generalized, and endpoint variables are not invertible and are well defined only in the direction of the arrows. For example, given the angular configuration of the arm, q, it is possible to derive the position of the hand, r. However the same position of the hand can be reached with different joint angles. Similarly, given the forces exerted by all the muscles on a joint, one derives the net torque while any value of torque can be obtained with an infinite number of muscle force combinations. How is it possible for the CNS to map a desired movement plan, in terms of endpoint behavior, into a corresponding command for the muscles? As shown in figure 37.4*B*, force fields provide additional pathways that map motions into forces. This shows how mechanics simplifies computations that would be otherwise unmanageable and ill posed. Thus, for example, a muscle synergy determines the force generated by the muscles in response to a stretch applied at any operating length. When multiple synergies are induced by a pattern of motor commands (u in figure 37.4*B*), their net effect is a field of torque vectors: For each configuration and state of motion of the joints, there is one and only one corresponding torque vector.

The planning of a desired behavior can in turn be expressed as a force field that maps the discrepancy between actual and desired state of the endpoint into a corrective force. In this way, a field of forces that converge on the target and that vanish there provides a detailed specification for the task of reaching a target with the hand. If an obstacle is interposed along the hand path, the concurrent goal of avoiding a collision can be represented as a field of forces that diverge from the obstacle. This mathematical representation has been proven successful in dealing with problems

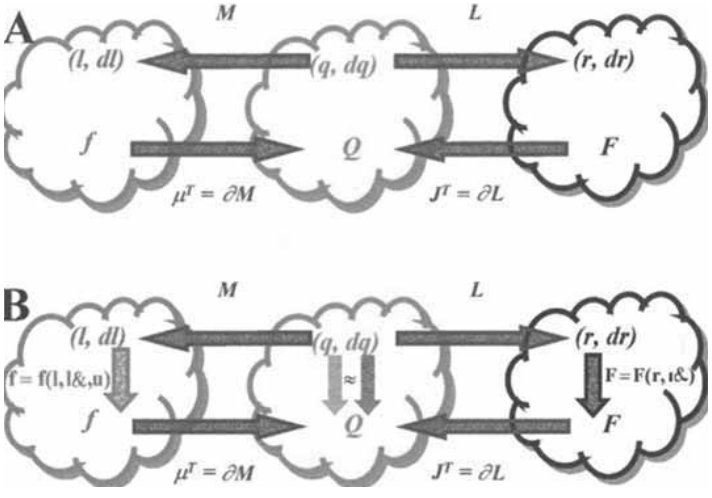


FIGURE 37.4 Coordinate transformations for planning and control of movement in a redundant limb. (*A*) Kinematic and force transformations for the human arm between muscle coordinates, joint coordinates, and endpoint coordinates. Arrows indicate the directions in which the transformations are well posed. Abbreviations: l, muscle lengths; q, joint angles; r, hand position; F, hand force; Q, joint torque; f, muscle force. M represents the transformation from joint angles to muscle lengths and L the transformation from joint angles to hand position αM and αL are the respective Jacobian matrices. (*B*) The vertical arrows map motion variables (position and velocity) onto force variables. They represent force fields. On the left is the force field generated by the muscles. On the right is the force field in endpoint coordinates that represents a desired behavior. Both the endpoint field and the muscle field have a well-defined image in joint coordinates, and the implementation of a desired behavior can be represented as a problem of approximation.

of robot motion planning (Khatib, 1986; Rimon & Koditschek, 1989). Although a literal implementation of this approach within the CNS may seem unlikely, one should observe that when we plan an action, such as reaching for a target, we are specifying not only a point of space but also additional requirements, such as those of remaining at rest at the target and avoiding collisions. Force fields provide a rigorous framework for expressing these concurrent demands by exploiting a mechanism of superposition. It is critical to observe that once a force field is given in endpoint coordinates, it is always possible to translate this field in joint coordinates despite the kinematic redundancy of the arm. The arrows on the right-hand side of figure 37.4*B* provide a path for this transformation: (1) The joint angles and angular displacements are mapped via the direct kinematics into a corresponding position and displacement of the hand; (2) the planned force field assigns a hand force to this position and displacement of the hand; and (3) the hand force is mapped into the corresponding joint torque by the (direct) Jacobian matrix of the arm.

It is apparent from the above considerations that the fields associated with muscle synergies and the fields associated with the description of a task find a common representation in joint coordinates (more generally, in generalized coordinates). In this common geometrical space, plans can be mapped into action by finding the combination of synergies that generates the best approximation of the planned field (*dashed arrows* in figure 37.4*B*). This is a process that does not require ill-posed inversions of redundant maps. Most important, force fields provide a way to compose building blocks of planning and building blocks of control through a single straightforward rule of linear summation.

Slotine and Lohmiller (2001) have developed a theoretical analysis of biological and robotic motor control based on the general concept of contraction dynamics (Lohmiller & Slotine, 1998). They pointed out that the mechanisms of biological evolution tend to favor behaviors that are not only successful but also, and most important, stable. The requirement of stability imposes a strong constraint on motor primitives, such as muscle synergies: Successful motor primitives should produce stable behaviors when they are acting individually and also when they operate in combination with other primitives. Convergent force fields such as those generated by spinal circuits satisfy this fundamental requirement. The modularity of control established by these fields has the critical property of ensuring stability to motor behaviors, when individual muscle synergies operate in isolation as well as when they operate in combination with other synergies, resulting in the summation of the respective force fields.

This view goes beyond the execution of a preplanned trajectory, as it allows for uncertainty in the knowledge of the environment and of the limb's mechanical properties and for a broader concept of motor planning that includes not only the generation of movements, but also the exertion of contact forces. As for the first point, the uncertainty, one should observe that a force field does not specify the accurate requirement of a position or of a trajectory but rather specifies the tradeoff between a desired position (or trajectory) and the force that is exerted in response to a deviation from this desired trajectory.

Optimal control and uncertainty: A computational basis for motor synergies

Once one accepts the general notions that muscles are organized into synergies and that synergies are combined by the CNS to form a repertoire of movements from a simple combinatorial "syntax" (Flash & Hochner, 2005), the next question to address is how these synergies are constructed from a large palette of possible muscle combinations. How does the motor system choose the particular patterns of stereotypical neuromusculat activations that constitute the natural vocabulary of motor primitives? To address this question, Todorov and coworkers (Todorov & Jordan, 2002; Todorov, Li, & Xiuchuan, 2005) have proposed that patterns of motor control arise from the goal of making movements as efficient as possible, given their goal and the constraints in which movements take place. An important constraint arises from motor noise.

The analysis of motor unit activities (Matthews, 1996) suggests a major role of synaptic noise in the excitation of motor neurons. Harris and Wolpert (1998) have proposed that the observed smoothness of natural motions that is observed in different motor behaviors (arm and eye movements, for example) may be accounted for by assuming that the biological controller minimizes the final error while being subject to signal dependent noise. This proposal is based on the idea that violations of smoothness, such as a large swing in a trajectory, are associated with large-amplitude control signals. Given that the signal variance accumulates additively along a movement, the net expected outcome of a jerky motion is a larger variance at the final point. Similar considerations are at the basis of a more general framework that was proposed recently by Todorov and Jordan (2002). They observed that in the presence of redundancy, one may identify within the space of control signals a lower-dimensional "task-relevant" manifold. This contains the combinations of motor commands that have a direct impact on the achievement of the established goal. Because of redundancy, at each point of this manifold, there is a "null space" of control signals that do not affect the execution of the task. For example, when we place the index finger on a letter key, we may do so with an infinite number of arm configurations. A common observation across a variety of behaviors is that variability tends

to be higher in the task-irrelevant dimensions. Todorov and Jordan (2002) consider this to be a direct consequence of optimal feedback control. According to this scheme, the control system aims to minimize the expected error on the final target in the presence of signal-dependent noise. While the outcomes of the optimization may depend upon the specific distribution of variability among the system of actuators, the simulations presented by these authors indicate a general tendency of the control system to place the highest variance in the task-irrelevant dimensions so as to achieve a higher degree of precision in the task-relevant dimensions. This view of the biological control system brings about two important (although yet to be proven valid) concepts:

1. that the control system is not necessarily concerned with the explicit planning of movement trajectories but rather is concerned with the attainment of final goals with the least amount of variance and
2. that the space spanned by the task-irrelevant dimensions plays the role of a "variance buffer," where the noise generated by the control signals has the largest effect so as to attain a higher performance in the space defined by the task.

Although this is a promising approach, with potentially important implications for the design of biomimetic controllers, the evidence for the explicit planning of trajectories remains rather strong (Dingwell, Mah, & Mussa-Ivaldi, 2002; Mosier, Scheidt, Acosta, & Mussa-Ivaldi, 2005; Flash, 1987; Shadmehr & Mussa-Ivaldi, 1994). It is difficult to reconcile the observation of regularities observed in endpoint coordinates, such as the execution of smooth rectilinear motions of the hand, with properties such as signal-dependent noise that concern the behavior of muscles and joints. To see this, consider a movement of the hand between two targets and suppose that the noise introduced by the shoulder muscles is greater than the noise introduced by the elbow muscles. If the only goal of the controller were to reduce the error on the final target, then the trajectory would be chosen to minimize the activity of the shoulder muscles. Obviously, a different trajectory would be chosen if the elbow muscles were the main source of noise. This prediction appears difficult to reconcile with the observation of smooth and quasi-rectilinear hand paths. Nevertheless, the framework of optimal control is important, as it establishes a direct one-to-one relationship between the goals of an action and the pattern of activations that are best suited to attain this goal. This approach provides "normative models" of motor control (Kording, 2007), which predict behavior based on the optimization of a cost function. This approach can also be used "in reverse" by searching for optimization functions, given a desired behavior. However, this is often seen as a weakness of optimization-based theories, as they may be fine-tuned to "explain" any observed pattern of behavior. The use of the optimality principle has a long and distinguished history in motor control (Hogan, 1984; Stein et al., 1986; van Beers, Baradua, & Wolpert, 2002). However, more recently, optimal control theory has been considered a promising framework for identifying patterns of meaningful muscle synergies. Chhabra and Jacobs (2006) have used Todorov's optimal feedback control to derive optimal patterns of joint torques for generating a large repertoire of reaching movements by a simulated arm. Then, as the joint torques are assumed to be produced by antagonist pairs of muscles acting across the shoulder and elbow joint, Chhabra and Jacobs used the nonnegative matrix factorization of d'Avella and colleagues (2003) to identify a set of muscle synergies that are competent to generate the optimal control torques. While this approach provides some general insight on the features of the motor synergies that may arise from optimal control, it is limited by the lack of available knowledge about the relationship between neuromuscular activities and their mechanical outcomes in terms of force. Therefore motor synergies expressed in terms of muscle forces are not easily translated into motor synergies in terms of neural or EMG activities, as in the work of d'Avella and colleagues (2003).

Todorov's approach to motor synergies (Todorov et al., 2005) is based on stochastic optimal control theory and may not yield an immediate interpretation of neural patterns. However, it has the potential to derive motor synergies based on the combined goals of satisfying optimal performance criteria and of enforcing compositionality. This is possible because the most general solution of optimization problems is obtained—in the continuous domain—by solving a partial differential equation, the Hamilton-Jacobi-Bellman (HJB) equation (Kirk, 1970), which is analogous to the Hamilton-Jacobi equation of classical mechanics and to the Schrödinger equation of quantum mechanics (Goldstein, 1980). The latter allowed deriving atomic structures as superposition of waves, a very primal form of compositionality. Todorov (2006) has recently obtained a similar result for optimal control. He derived a linear formulation of the HJB equation that allows obtaining primitives of control from a system of basis eigenfunctions of a linear operator. In this case, as for the wave functions of Schrödinger, any combination of optimal control policies still satisfies the same optimality condition as the individual policies.

Although a large collection of open issues remains, the combination of computational models based on nonlinear control and optimization theories with more advanced techniques for extracting information from neuromuscular patterns yields the promise of understanding how biological motor behavior is shaped by the interaction of evolutionary process and day-to-day learning mechanisms.

REFERENCES

Abend, W. K., Bizzi, E., & Morasso, P. (1982). Human arm trajectory formation. *Brain*, *105*, 331–348.

Arbib, M. A. (1981). Perceptual structures and distributed motor control. In V. B. Brooks (Ed.), *Handbook of physiology, Section 2: The nervous system, motor control, Part 1* (vol. 2, pp. 1449–1480). Bethesda, MD: American Physiological Society.

Bizzi, E., Accornero, N., Chapple, W., & Hogan, N. (1984). Posture control and trajectory formation during arm movement. *J. Neurosci.*, *4*, 2738–2744.

Bizzi, E., Mussa-Ivaldi, F. A., & Giszter, S. F. (1991). Computations underlying the execution of movement: A biological perspective. *Science*, *253*, 287–291.

Bizzi, E., Polit, A., & Morasso, P. (1976). Mechanisms underlying achievement of final position. *J. Neurophysiol.*, *39*, 435–444.

Bishop, R. L., & Goldberg, S. I. (1980). *Tensor analysis on manifolds*. New York: Dover.

Chhabra, M., & Jacobs, R. A. (2006). Properties of synergies arising from a theory of optimal motor behavior. *Neural Comput.*, *18*, 2320–2342.

Cvitanovic, P. (2000). Chaotic field theory: A sketch. *Physics A*, *288*, 61–69.

d'Avella, A., & Bizzi, E. (2005). Shared and specific muscle synergies in natural motor behaviors. *Proc. Natl. Acad. Sci. USA*, *102*(8), 3076–3081.

d'Avella, A., Portone, A., Fernandez, L., & Lacquaniti, F. (2006). Control of fast-reaching movements by muscle synergy combinations. *J. Neurosci.*, *26*(30), 7791–7810.

d'Avella, A., Saltiel, P., & Bizzi, E. (2003). Combinations of muscle synergies in the construction of a natural motor behavior. *Nat. Neurosci.*, *6*(3), 300–308.

d'Avella, A., & Tresch, M. C. (2002). Modularity in the motor system: Decomposition of muscle patterns as combinations of time-varying synergies. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems* (vol. 14). Cambridge, MA: MIT Press.

Davidson, P. R., & Wolpert, D. M. (2004). Internal models underlying grasp can be additively combined. *Exp. Brain Res.*, *155*(3), 334–340.

Dingwell, J. B., Mah, C. D., & Mussa-Ivaldi, F. A. (2002) Manipulating objects with internal degrees of freedom: Evidence for model-based control. *J. Neurophysiol.*, *88*(1), 222–223.

Feldman, A. G. (1966). Functional tuning of the nervous system with control of movement or maintenance of steady posture: II. Controllable parameters of the muscles. *Biophysics*, *11*, 565–578.

Feldman, A. G., & Latash, M. L. (2005). Testing hypotheses and the advancement of science: Recent attempts to falsify the equilibrium point hypothesis. *Exp. Brain Res.*, *161*(1), 91–103.

Flanagan, J. R., & Wing, A. M. (1997). The role of internal models in motion planning and control: Evidence from grip force adjustments during movements of hand-held loads. *J. Neurosci.*, *17*(4), 1519–1528.

Flash, T. (1987). The control of hand equilibrium trajectories in multi-joint arm movements. *Biol. Cybern.*, *57*(4–5), 257–274.

Flash, T., & Hochner, B. (2005). Motor primitives in vertebrates and invertebrates. *Curr. Opin. Neurobiol.*, *15*(6): 660–666.

Flash, T., & Hogan, N. (1985). The coordination of arm movements. An experimentally confirmed mathematical model. *J. Neurosci.*, *5*, 1688–1703.

Gister, S. F., & Kargo, W. J. (2000). Conserved temporal dynamics and vector superposition of primitives in frog wiping reflexes during spontaneous extensor deletions. *Neurocomputing*, *32–33*, 775–783.

Giszter, S. F., Mussa-Ivaldi, F. A., & Bizzi, E. (1993). Convergent force fields organized in the frog's spinal cord. *J. Neurosci.*, *13*, 467–491.

Goldstein, H. (1980). *Classical mechanics* (2nd ed.). Reading, MA: Addison-Wesley.

Grillner, S. (1981). Control of locomotion in bipeds, tetrapods, and fish. In V. B. Brooks (Ed.), *Handbook of physiology* (vol. 2, pp. 1179–1236). Bethesda, MD: American Physiological Society.

Grillner, S. (1985). Neurobiological bases of rhythmic motor acts in vertebrates. *Science*, *228*, 143–149.

Grillner, S., & Zangger, P. (1979). On the central generation of locomotion in the low spinal cat. *Exp. Brain Res.*, *34*, 241–261.

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784.

Hogan, N. (1984). An organizing principle for a class of voluntary movement control. *J. Neurosci.*, *4*, 2745–2754.

Holdefer, R. N., & Miller, L. E. (2002). Primary motor cortical neurons encode functional muscle synergies. *Exp. Brain Res.*, *146*(2), 233–243.

Jose, J. V., & Saletan, E. J. (1998). *Classical dynamics: A contemporary approach*. Cambridge, UK: Cambridge University Press.

Kargo, W. J., & Giszter, S. F. (2000). Rapid correction of aimed movements by summation of force-field primitives. *J. Neuorosci.*, *20*, 409–426.

Kawato, M., & Wolpert, D. (1998). Internal models for motor control. *Novartis Found Symp.*, *218*, 291–304; discussion, 304–307.

Khatib, O. (1986). Real-time obstacle avoidance for manipulators and mobile robots. *Int. J. Robotics Res.*, *5*, 90–99.

Kirk, D. E. (1970). *Optimal control theory*. Englewood Cliffs, NJ: Prentice Hall.

Kording, K. (2007). Decision theory: What "should" the nervous system do? *Science, 318*, 606–610.

Krakauer, J. W., Ghilardi, M. F., & Ghez, C. (1999). Independent learning of internal models for kinematic and dynamic control of reaching. *Nat. Neurosci.*, *2*(11), 1026–1031.

Krouchev, N., Kalaska, J. F, & Drew, T. (2006). Sequential activation of muscle synergies during locomotion in the intact cat as revealed by cluster analysis and direct decomposition. *J. Neurophysiol.*, *96*(4), 1991–2010.

Lemay, M. A. Galagan, J. E. Hogan, N., & Bizzi, E. (2001). Modulation and vectorial summation of the spinalized frog's hindlimb end-point force produced by intraspinal electrical stimulation of the cord. *IEEE Trans. Neural Syst. Rehab. Eng.*, *9*(1), 12–23.

Lohmiller, W., & Slotine, J.-J. E. (1998). On contraction analysis for nonlinear systems. *Automatica*, *34*, 683–696.

Macpherson, J. M. (1991). How flexible are muscle synergies? In D. R. Humphrey & H. J. Freund (Ed.), *Motor control: Concepts and issues* (pp. 33–47). New York: John Wiley.

Matthews, P. B. (1996). Relationship of firing intervals of human motor units to the trajectory of post-spike after-hyperpolarization and synaptic noise. *J. Phyisiol.*, *492*, 597–682.

Morasso, P. (1981). Spatial control of arm movements. *Exp. Brain Res.*, *42*, 223–227.

Mosier, K. M., Scheidt, R. A., Acosta, S., & Mussa-Ivaldi, F. A. (2005). Remapping hand movements in a novel geometrical environment. *J. Neurophysiol.*, *94*(6), 4362–4372.

Mussa-Ivaldi, F. A., Giszter, S. F., & Bizzi, E. (1994). Motor learning through the combination of primitives. *Proc. Natl. Acad. Sci. USA*, *91*, 7534–7538.

Rimon, E., & Koditschek, D. E. (1989). The construction of analytic diffeomorphisms for exact robot navigation on star worlds. In *Proceedings of the 1989 IEEE International Conference on Robotics and Automation* (pp. 21–26). Los Alamitos, CA: IEEE Computer Society Press.

Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Englewood Cliffs, NJ: Prentice Hall.

Sabes, P. N. (2000). The planning and control of reaching movements. *Curr. Opin. Neurobiol.*, *10*(6), 740–746.

Saltiel, P., Wyler-Duda, K., d'Avella, A., Tresch, M. C., & Bizzi, E. (2001). Muscle synergies encoded within the spinal cord: Evidence from focal intraspinal NMDA iontophoresis in the frog. *J. Neurophysiol.*, *85*(2), 605–619.

Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.*, *14*(5, Pt 2), 3208–3224.

Slotine, J. J., & Lohmiller, W. (2001). Modularity, evolution, and the binding problem: A view from stability theory. *Neural Net.*, *14*(2), 137–145.

Soechting, J. F., & Lacquaniti, F. (1981). Invariant characteristics of a pointing movement in man. *J. Neurosci.*, *1*(7), 710–720.

Soechting, J. F., & Lacquaniti, F. (1989). An assessment of the existence of muscle synergies during load perturbations and intentional movements of the human arm. *Exp. Brain Res.*, *74*(3), 535–548.

Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.*, *14*(5, Pt 2), 3208–3224.

Spong, M. W., Hutchinson, S., & Vidyasagar, M. (2005). *Robot modeling and control*. Hoboken, NJ: John Wiley.

Stein, P. S., McCullough, M. L., & Currie, S. N. (1998). Spinal motor patterns in the turtle. *Ann. NY Acad. Sci.*, *16*, 142–154.

Stein, R. B., Oguztoreli, M. N., & Capaday, C. (1986). What is optimized in muscular movements? In N. L. Jones, N. McCartney, & A. J. McComas (Eds.), *Human muscle power* (pp. 131–150). Champaign, IL: Human Kinetics.

Thoroughman, K. A., & Taylor, J. A. (2005). Rapid reshaping of human motor generalization. *J. Neurosci.*, *25*, 8948–8953.

Ting, L. H., & Macpherson, J. M. (2005). A limited set of muscle synergies for force control during a postural task. *J. Neurophysiol.*, *93*, 609–613.

Todorov, E. (2005). Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Comput.*, *17*(5), 1084–108.

Todorov, E. (2006). *Primitives for optimal control*. Paper presented at Advances in Computational Motor Control V, Atlanta, GA. October 13, 2006.

Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.*, *5*(11), 1226–1235.

Todorov, E., Li, W., & Xiuchuan, P. (2005). From task parameters to motor synergies: A hierarchical framework for approximately optimal control of redundant manipulators. *J. Robot. Syst.*, 22, 691–710.

Tresch, M. C., Cheung, V. C. K., & d'Avella, A. (2006). Matrix factorization algorithms for the identification of muscle synergies: Evaluation on simulated and experimental data sets. *J. Neurophysiol.*, *95*, 2199–2212.

Tresch, M. C., Saltiel, P., & Bizzi, E. (1999). The construction of movement by the spinal cord. *Nat. Neurosci.*, *2*(2), 162–167.

Tresch, M. C., Saltiel, P., d'Avella, A., & Bizzi, E. (2002). Coordination and localization in spinal motor systems. *Brain Res. Brain Res. Rev.*, *40*(1–3), 66–79.

Tsetlin, M. L. (1973). *Automation theory and modeling of biological systems* (pp. 160–196). New York: Academic Press.

Uno, Y., Kawato, M., & Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement: Minimum torque-change model. *Biol. Cybern.*, *61*, 89–101.

van Beers, R. J., Baraduc, P., & Wolpert, D. M. (2002). Role of uncertainty in sensorimotor control. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *357*, 1137–1145.

Wolpert, D., Miall, R., & Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci.*, *2*, 338–347.

38 Basal Ganglia and Cerebellar Circuits with the Cerebral Cortex

RICHARD P. DUM AND PETER L. STRICK

ABSTRACT What are the functions of the basal ganglia and cerebellum? It is now clear that output of the basal ganglia and cerebellum targets motor, premotor, prefrontal, posterior parietal, and inferotemporal areas of cortex. These connections provide the basal ganglia and cerebellum with the anatomical substrate to influence not only the control of movement, but also many aspects of cognitive behavior like planning, working memory, sequential behavior, visuospatial perception, and attention. Similarly, abnormal activity in specific basal ganglia and cerebellar loops with the cerebral cortex may contribute to a variety of neuropsychiatric disorders, such as schizophrenia, autism, attention-deficit/hyperactivity disorder, and obsessive-compulsive disorder. Thus, defining the cortical targets of the basal ganglia and cerebellum provides important insights into their diverse motor and nonmotor function.

What are the functions of the basal ganglia and the cerebellum? Numerous reports describe the motor deficits associated with damage to these subcortical structures. As a consequence, concepts about basal ganglia and cerebellar function have focused primarily on their contributions to the generation and control of movement. We have used an anatomical approach to examine the macro-organization of basal ganglia and cerebellar connections with the cerebral cortex. In this chapter, we focus on one critical question: Which cortical areas are the target of the outputs from the basal ganglia and the cerebellum? The answers to this question lead to some novel and important insights about basal ganglia and cerebellar function.

Classically, the macro-organization of basal ganglia and cerebellar circuitry is described using a relatively simple hierarchical model. The "input layer" of basal ganglia processing is represented by the striatum (caudate, putamen, and ventral striatum). The functionally analogous level in cerebellar circuits is represented by specific pontine nuclei that send "mossy fiber" inputs to cerebellar cortex. A major source of afferents to the input layers of both circuits originates from widespread regions of the cerebral cortex, including motor, sensory, posterior parietal, prefrontal, cingulate, orbitofrontal, and temporal cortical areas. The "output layer" of basal ganglia processing is represented by the internal segment of the globus pallidus (GPi), the pars reticulata of the substantia nigra (SNpr), and the ventral pallidum. The comparable structures for cerebellar processing are the three deep cerebellar nuclei: dentate, interpositus, and fastigial. Neurons in the output layers of both circuits send their axons to the thalamus and, by this route, project back upon the cortex. Thus, a major structural feature of basal ganglia and cerebellar circuits is that they form loops with the cerebral cortex (e.g., Kemp & Powell, 1971; Allen & Tsukahara, 1974; Brooks & Thach, 1981). These loops were believed to function largely in the domain of motor control. Indeed, basal ganglia and cerebellar efferents were thought to terminate in a common region of the ventrolateral thalamus that projected largely to the primary motor cortex (M1). Thus these circuits were viewed as a neural substrate for enabling information from a diverse set of cortical areas to influence motor output at the level of M1. This view has been supported by the obvious motor symptoms that can result from basal ganglia and cerebellar dysfunction (for references and reviews, see Brooks & Thach, 1981; DeLong & Georgopoulos, 1981; Bhatia & Marsden, 1994).

Over the past 20 years, an accumulation of information about basal ganglia and cerebellar anatomy has led a number of investigators to challenge this view (e.g., Schell & Strick, 1984; Alexander, DeLong, & Strick, 1986; Goldman-Rakic & Selemon, 1990). It is now clear that basal ganglia and cerebellar efferents terminate in different subdivisions of the ventrolateral thalamus (for a review, see Percheron, François, Talbi, Yelnik, & Fénelon, 1996), which, in turn, project to a myriad of cortical areas. Thus the outputs from the basal ganglia and cerebellum influence more widespread regions of the cerebral cortex than was previously recognized.

On the basis of these and other anatomical results, Alexander and colleagues (1986) proposed that the basal ganglia participate in at least five separate loops with the cerebral cortex. These loops were based in part on their cortical target from the output layer of processing and were designated the skeletomotor, oculomotor, dorsolateral

RICHARD P. DUM Center for the Neural Basis of Cognition, Systems Neuroscience Institute and the Department of Neurobiology, University of Pittsburgh, Pittsburgh, Pennsylvania
PETER L. STRICK Veterans Affairs Medical Center; Center for the Neural Basis of Cognition, Systems Neuroscience Institute and the Department of Neurobiology, University of Pittsburgh, Pittsburgh, Pennsylvania

prefrontal, lateral orbitofrontal, and anterior cingulate circuits. According to this scheme, the output of the basal ganglia has the potential to influence not only the control of movement, but also higher-order cognitive and limbic functions that are subserved by prefrontal, orbitofrontal, and anterior cingulate cortex.

Similarly, Leiner, Leiner, and Dow (1986, 1991, 1993) suggested that cerebellar output is directed to prefrontal as well as motor areas of the cerebral cortex. They noted that in the course of hominid evolution, the lateral output nucleus of the cerebellum—the dentate—undergoes a marked expansion that parallels the expansion of cerebral cortex in the frontal lobe. They argued that the increase in the size of the dentate is accompanied by an increase in the extent of the cortical areas in the frontal lobe that are influenced by dentate output. As a consequence, Leiner and colleagues proposed that cerebellar function in humans has expanded to include involvement in certain language and cognitive tasks.

Attempts to test these proposals and map cerebellar and basal ganglia projections to the cerebral cortex have been hindered by a number of technical limitations. Chief among these are the multisynaptic nature of these pathways and the general inability of conventional tracers to label more than the direct inputs and outputs of an area. To overcome these and other problems, we developed the use of neurotropic viruses (herpes simplex virus type 1 [HSV1] and rabies virus) as transneuronal tracers in the central nervous system of primates (for references and review, see Strick & Card, 1992; Kelly & Strick, 2000, 2003). This tracing method can effectively label a chain of up to three synaptically linked neurons in a single experiment (Kelly & Strick, 2003, 2004). In this chapter, we will review some of the new observations that have come from using viruses to trace basal ganglia and cerebellar loops with the cerebral cortex. These observations have led to important insights about the cortical targets of these circuits and the functional domains they influence.

Primary motor cortex

Our first experiments used retrograde transneuronal transport of HSV1 to examine the organization of basal ganglia and cerebellar outputs to M1 (figure 38.1) (Hoover & Strick, 1993, 1999). We injected virus into physiologically identified portions of M1 (i.e., regions where face, arm, or leg movements were evoked by intracortical stimulation with currents < 25 μA). Then we set the survival time to allow transneu-

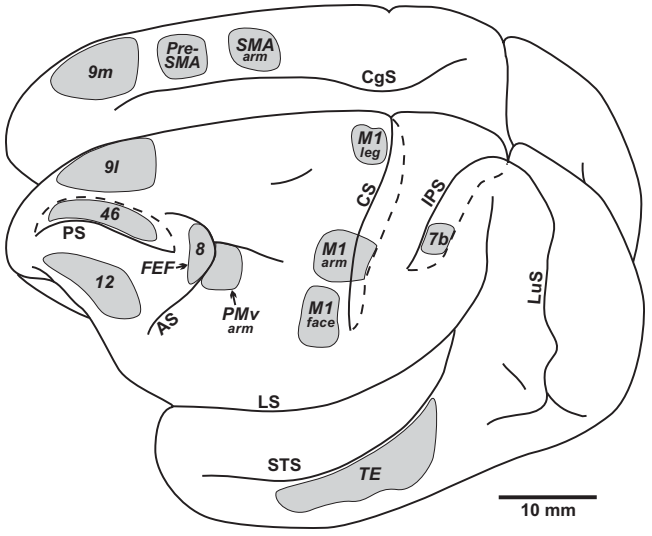


FIGURE 38.1 Virus injection sites in the cerebral cortex. The locations of virus injection sites (*shaded areas*) are shown on a view of the lateral surface and a mirror image of the medial wall of a cebus monkey brain. The numbers 7b, 8, 9m, 9l, 12, and 46 refer to cytoarchitectonic areas. Abbreviations: AS, arcuate sulcus; CS, central sulcus, FEF, frontal eye field; IPS, intraparietal sulcus, LS, lateral sulcus, LuS, lunate sulcus; M1, face, arm, and leg areas of the primary motor cortex; PMv_{arm}, arm area of the ventral premotor area; PreSMA, presupplementary motor area; PS, principal sulcus; SMA_{arm}, arm area of the supplementary motor area; STS, superior temporal sulcus; TE, area of inferotemporal cortex.

FIGURE 38.2 Origin of pallidal projections to M1, PMv, SMA, area 46, area 9. Representative coronal sections through the GPi of animals that received virus injections into different cortical areas (see figure 38.1). The *dots* indicate the positions of neurons labeled by retrograde transneuronal transport of virus. The maps display labeled neurons found on two or three adjacent sections whose approximate anterior-posterior location is indicated at the bottom of each section outline. Abbreviations: GPe, external segment of globus pallidus; GPi, internal segment of the globus pallidus; o, outer portion of the interal segment of globus pallidus; i, inner portion of the internal segment of globus pallidus. (Adapted from Middleton & Strick, 1996b.)

ronal transport of virus to label "second-order" neurons that are the origin of basal ganglia and cerebellothalamocortical inputs to M1. The brain of each animal was processed by using immunohistochemical procedures to demonstrate the location of virus-specific antigen in infected neurons (Strick & Card, 1992; Kelly & Strick, 2000).

Three major results came from these experiments. First, we found that M1 is richly innervated by the output nuclei of the basal ganglia and cerebellum. The densest projections originate from GPi (figures 38.2 and 38.3, M1 arm) and the dentate (figures 38.4 and 38.5, M1 arm). Less dense projections originate from portions of the SNpr and interpositus. Second, we found that both the GPi and the dentate are somatotopically organized with separate face, arm, and leg areas that project via the thalamus to the face, arm, and leg areas of M1. Third, and perhaps most surprising, we discovered that projections to M1 originate from only 15% of the volume of the GPi and about 30% of the volume of the dentate. Thus, the output to M1 originates from restricted portions of each subcortical nucleus. This result implies that the majority of the output from the basal ganglia and cerebellum is directed to other cortical areas.

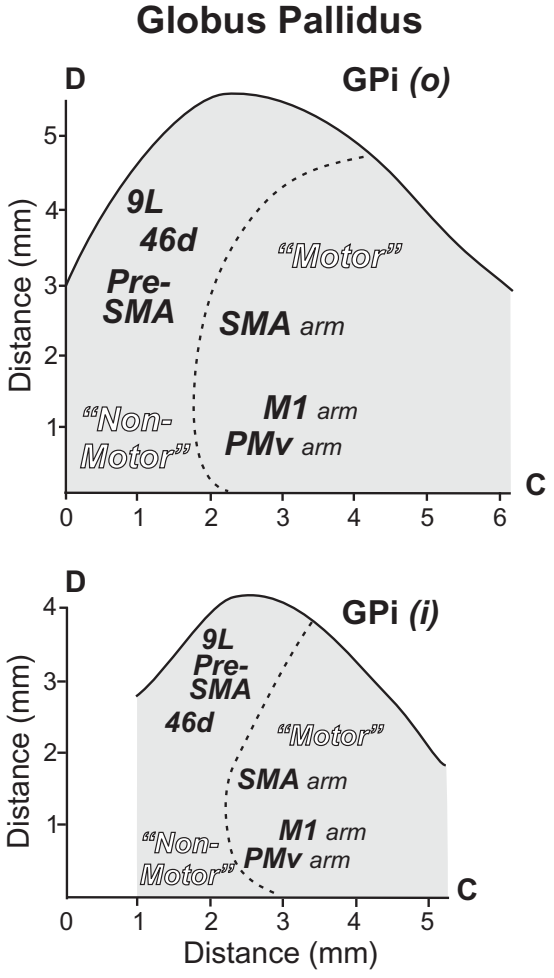


FIGURE 38.3 Summary map of the basal ganglia output channels. The outer and inner segments of the GPi are shown as separate unfolded maps (for details of unfolding, see Akkal et al., 2007). This map provides a planar view of the rostrocaudal and dorsoventral location of output channels in each segment of the GPi. The cortical target of each output channel is placed at the site of the peak labeling following retrograde transneuronal transport of virus from that cortical area. Note that the GPi can be divided into "motor" and "nonmotor" domains based on the grouping of the cortical targets of its output channels. Abbreviations: D, dorsal; C, caudal; see also figures 38.1 and 38.2. (Adapted from Akkal et al., 2007.)

Premotor areas

Our next experiments used virus tracing to examine basal ganglia and cerebellar projections to the arm representations of premotor areas in the frontal lobe (figure 38.1) (Hoover & Strick, 1993; Akkal, Dum, & Strick, 2007). Injections of virus into either the ventral premotor area (PMv) or the supplementary motor area (SMA) consistently labeled neurons in the middle of the GPi rostrocaudally. Within this region, neurons labeled after injections into the SMA, M1, or PMv formed separate clusters in a dorsal-to-ventral arrangement (figures 38.2 and 38.3). These observations indicate that pallidal output is not confined to M1 but projects via the thalamus to multiple premotor areas in the frontal lobe (see also Jinnai, Nambu, Tanibuch, & Yoshida, 1993; Inase & Tanji, 1995; Sakai, Inase, & Tanji, 1999). Furthermore, the arm representation of each motor area receives input from a topographically distinct set of GPi neurons. We have proposed that this arrangement creates distinct "output channels" in the sensorimotor portion of GPi (Hoover & Strick, 1993; Akkal et al., 2007).

We found a similar topographic organization of output neurons in the dentate. Injections of virus into the arm representations of M1, PMv, and SMA labeled clusters of neurons in the middle of the dentate rostrocaudally (figures 38.4 and 38.5) (Middleton & Strick, 1997; Akkal et al., 2007). However, the "hotspot" of each cluster appeared to be centered in a slightly different region of the dentate. The hotspots for the different motor areas are shown on a single unfolded map of the dentate (figure 38.5). This diagram emphasizes two important observations. First, the arm representations of the PMv and SMA are the target of output from the dentate. Second, the output channels to the different arm representations are clustered together in a common region of the dorsal dentate. This observation raises the possibility that the dorsal dentate contains a single integrated map of the body in which the maps for output channels to different cortical areas are in register within the nucleus. In any event, the dentate, like the GPi, contains distinct output channels that innervate different cortical motor areas.

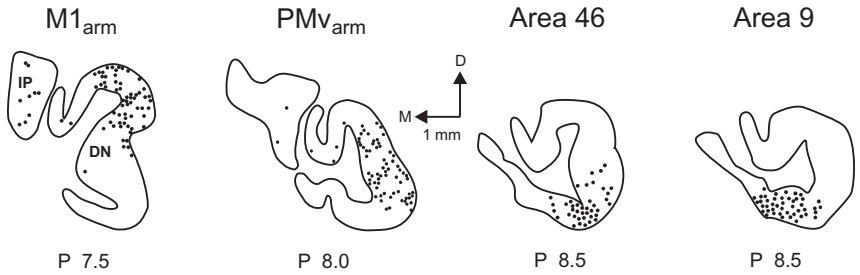


FIGURE 38.4 Origin of cerebellar projections to M1, PMv, area 46, and area 9. Representative coronal sections through the dentate and interpositus nuclei of animals that received virus injections into different cortical areas (see figure 38.1). Conventions are according to figure 38.2. Abbreviations: D, dorsal; DN, dentate nucleus; IP, interpositus nucleus; M, medial. (Adapted from Middleton & Strick, 1996b.)

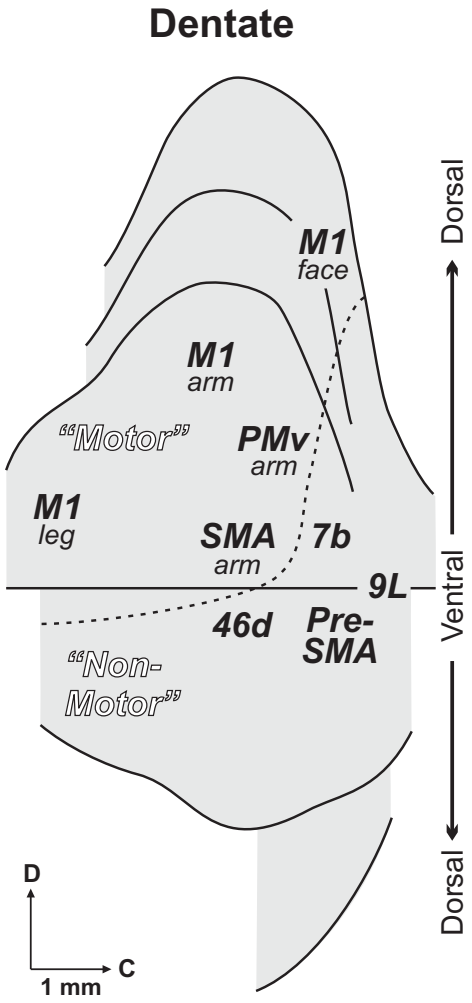


FIGURE 38.5 Summary map of dentate output channels. The dentate is displayed as an unfolded map (for details of unfolding, see Dum & Strick, 2003). The cortical target of each output channel is placed at the site of the peak labeling following retrograde transneuronal transport of virus from that cortical area. Note that the dentate can be divided into "motor" and "nonmotor" domains based on the grouping of the cortical targets of its output channels. Abbreviations as in figure 38.1. (Adapted from Dum & Strick, 2003; Akkal et al., 2007.)

Results from single-neuron recording experiments in awake trained monkeys provide physiological support for the existence of distinct output channels in the GPi and dentate (Mushiake & Strick, 1993, 1995; Strick, Dum, & Picard, 1995). These studies suggest that individual output channels are involved in different aspects of motor behavior. Specifically, some output channels appear to be especially concerned with movements that are internally generated, whereas others appear to be devoted to movements guided by exteroceptive cues. Taken together, these observations indicate that the basal ganglia and cerebellum have the capacity to influence a broad range of motor behavior using output channels that project to the premotor areas in the frontal lobe as well as to M1. Thus, the skeletomotor circuit of Alexander and colleagues (1986) is more accurately viewed as multiple discrete channels to each of the cortical motor areas (figure 38.6). A similar arrangement of output channels characterizes skeletomotor output from the dentate.

Frontal eye field

We have also used transneuronal transport of virus to examine subcortical inputs to the frontal eye field (FEF) (figure 38.1) (Lynch, Hoover, & Strick, 1994). The results of prior studies with conventional tracers led to the proposal that the FEF receives input via the thalamus from three major subcortical nuclei: SNpr, the superior colliculus (SC), and the deep cerebellar nuclei. To test this proposal, we injected virus into physiologically identified portions of the FEF (i.e., regions where eye movements were evoked by intracortical stimulation with currents $< 50\ \mu A$). Neurons labeled by retrograde transneuronal transport were found in lateral portions of SNpr, the optic and intermediate gray layers of the SC, and ventrally in the caudal third of the dentate nucleus. Within the dentate, labeled neurons were confined to its posterior pole, where some neurons exhibit activity correlated with saccadic eye movements (van Kan, Houk, & Gibson, 1993). Within the basal ganglia, FEF injections labeled neurons in a posterior and lateral region of

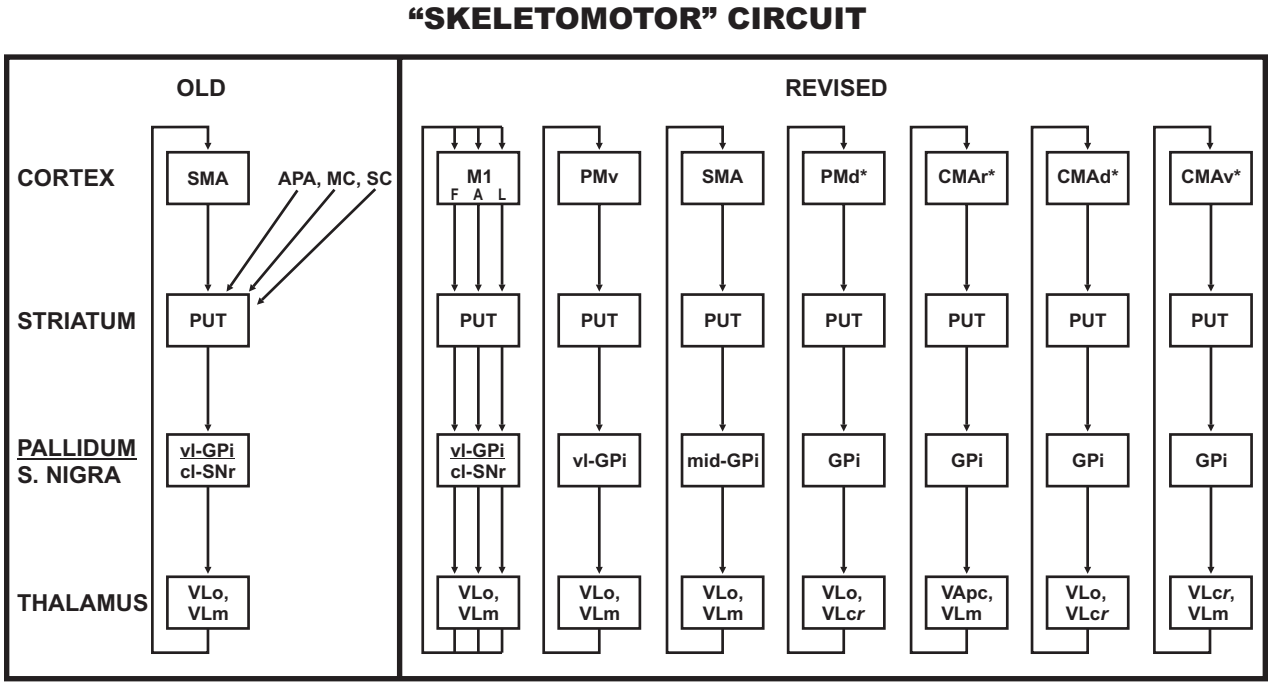


FIGURE 38.6 The original skeletomotor circuit proposed by Alexander, DeLong, and Strick (1986) and our revised scheme. *Asterisks* indicate loops whose existence is suspected but not specifically tested using virus transport. Cortical abbreviations: CMAd, dorsal cingulate motor area; CMAr, rostral cingulate motor area; CMAv, ventral cingulate motor area; M1, primary motor cortex; PMd, dorsal premotor area; PMv, ventral premotor area; SMA, supplementary motor area. Basal ganglia abbreviations: GPi, internal segment of globus pallidus; PUT, putamen; SNr substantia nigra pars reticulata; cl, caudolateral; mid, middle; vl, ventrolatedal. Thalamic abbreviations: VApc, nucleus ventralis anterior, parvocellular portion; VLcc, nucleus ventralis lateralis pars caudalis, caudal division; VLcr, nucleus ventralis lateralis pars caudalis, rostral division; VLm, nucleus ventralis lateralis pars medialis; VLo, nucleus ventralis lateralis pars oralis. (Adapted from Middleton & Strick, 2000.)

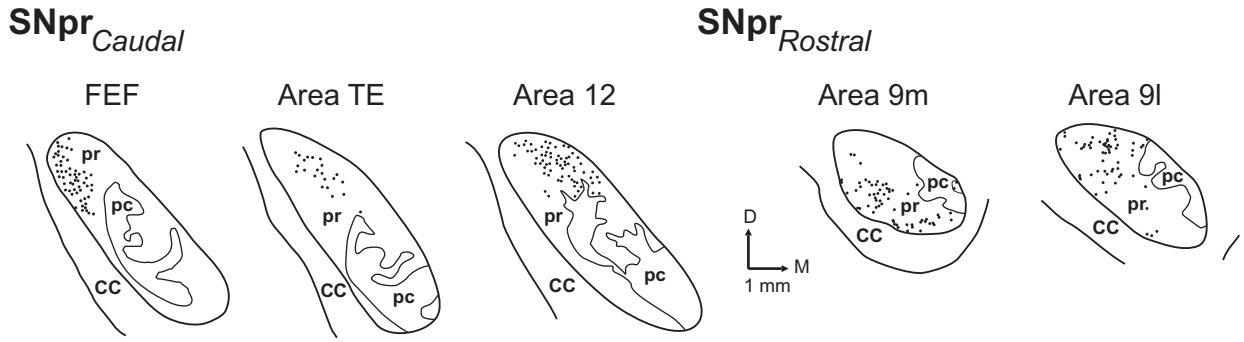


FIGURE 38.7 Origin of nigral projections to the FEF, area TE, area 12, area 9m, and area 9l. Coronal sections indicating the location of labeled neurons in the caudal and rostral regions of the SNpr following virus injections into the different cortical areas (see figure 38.1). Abbreviations: CC, crus cerebri; pc, pars compacta; pr, pars reticulata. (Adapted from Middleton & Strick, 1996b.)

SNpr (figure 38.7, FEF) where neurons also display changes in activity related to saccadic eye movements (Hikosaka & Wurtz, 1983a, 1983b). Overall, the regions of the basal ganglia and cerebellum that were labeled after FEF injections of virus were strikingly different from those labeled after injections into any of the skeletomotor areas of the frontal lobe. Thus the output channels in the basal ganglia and cerebellum that are concerned with oculomotor function are distinct from those concerned with skeletomotor function.

Prefrontal cortex

It is clear from the studies reviewed above that the output nuclei of the basal ganglia and cerebellum have well-organized projections to skeletomotor and oculomotor areas

of cortex. Nevertheless, substantial portions of these output nuclei do not project to cortical motor areas. This observation raises the possibility that the remaining portions of these output nuclei target nonmotor areas of cortex. Because of prior suggestions that the basal ganglia and cerebellum influence some of the cognitive operations that are normally thought to be subserved by the frontal lobe (Alexander et al., 1986; Leiner et al., 1986, 1991, 1993), we used virus tracing to test whether basal ganglia and cerebellar projections to prefrontal cortex provide an anatomical substrate for this influence.

Our experiments focused on subfields within areas 9, 12, and 46 of the prefrontal cortex (Middleton & Strick, 1994, 2001, 2002). Each of these areas appears to be involved in aspects of "working memory" and is thought to guide behavior based on transiently stored information rather than immediate external cues (for reviews, see Passingham, 1993; Goldman-Rakic, 1996; Fuster, 1997). Virus injections into area 9, 12, or 46 labeled many neurons in the output nuclei of the basal ganglia (figures 38.2, 38.3, and 38.7). Injections into area 12 labeled neurons in a localized portion of SNpr. In contrast, injections into area 46 labeled neurons largely in GPi. Area 9 injections labeled neurons in both the SNpr and GPi. The topographic nature of basal ganglia projections to prefrontal cortex is further emphasized by the finding that different regions within the rostral SNpr project to medial and lateral portions of area 9 (figure 38.7, areas 9m and 9l). In all cases, the locations of the neurons labeled in the GPi and SNpr after injections into prefrontal areas of cortex are different from the locations of neurons labeled after injections into motor areas of cortex.

Virus injections into areas 9 and 46 (but not area 12) labeled neurons in ventral regions of the dentate nucleus (figures 38.4 and 38.5). The neurons that were labeled after injections into area 9 were found largely medial and caudal to those labeled by injections into area 46. The ventral regions of the dentate that project to these nonmotor areas in the frontal lobe clearly differ from the more dorsal regions of this nucleus that innervate motor areas of the cortex (figures 38.4 and 38.5). Thus, both the basal ganglia and the cerebellum project via the thalamus to multiple areas of prefrontal cortex. Moreover, the output channels in the basal ganglia and cerebellum that influence prefrontal areas of cortex are separate from those that influence motor areas of cortex. This observation suggests that GPi and the dentate can be divided into motor and nonmotor domains (figures 38.3 and 38.5) (Dum & Strick, 2003; Akkal et al., 2007).

Although the presupplementary motor area (PreSMA) has traditionally been included with the motor areas of the frontal lobe, a number of recent observations emphasize the nonmotor nature of this cortical area (for a review, see Picard & Strick, 2001). For example, unlike the cortical motor areas, the PreSMA does not project directly to M1 or to the spinal cord. Instead, the PreSMA is densely interconnected with regions of prefrontal cortex. We used virus tracing to test whether basal ganglia and cerebellar projections to the PreSMA originate from the motor or nonmotor domains of the GPi and the dentate (Akkal et al., 2007). We found that the output channel in the GPi that projects to the PreSMA is located dorsally in a rostral portion of the nucleus (figure 38.3). The output channel in the dentate that projects to the PreSMA is located in a ventral part of the nucleus (figure 38.5). Thus the output channels to the PreSMA in both the GPi and the dentate are adjacent to output channels that project to regions of prefrontal cortex rather than near output channels to the cortical motor areas (figures 38.3 and 38.5). These observations provide further support for the proposal that the PreSMA is more similar to regions of prefrontal cortex than it is to a cortical motor area (Picard & Strick, 2001; Akkal et al., 2007).

Posterior parietal cortex

Areas 5 and 7 in the posterior parietal cortex are known to project to the input stage of basal ganglia and cerebellar processing (e.g., Kemp & Powell, 1971; Glickstein, May, & Mercier, 1985; Cavada & Goldman-Rakic, 1991; Yeterian & Pandya, 1993; Schmahmann & Pandya, 1997). These connections led us to ask whether the posterior parietal cortex is a target of basal ganglia and cerebellar output (Clower, West, Lynch, & Strick, 2001; Clower, Dum, & Strick, 2005). Our results demonstrate that a portion of area 7b in the intraparietal sulcus is the target of output from the dentate nucleus, whereas a portion of area 7b on the cortical surface is the target of output from the SNpr as well as from the dentate nucleus. These results clearly indicate that the sphere of influence of basal ganglia and cerebellar output extends to include portions of the posterior parietal cortex. Space limitations do not allow us to describe the full implications of basal ganglia and cerebellar projections to posterior parietal cortex. Instead, we will highlight two specific proposals about these circuits. We have suggested that the cerebellar projection to the posterior parietal cortex may provide signals that contribute to the sensory recalibration that occurs during some adaptation paradigms (Clower et al., 2001). On the other hand, we have suggested (Clower et al., 2005) that abnormal signals in the basal ganglia projection to the posterior parietal cortex may contribute to the visuospatial deficits that are observed in some patients with basal ganglia lesions (Karnath, Himmelbach, & Rorden, 2002).

Inferotemporal cortex

In general, each of the cortical areas found to receive input from the basal ganglia or cerebellum is known to send pro-

jections back to these subcortical nuclei. This anatomical arrangement suggests that many cortical areas in the frontal lobe participate in "closed loop" circuits with the basal ganglia and cerebellum. To test whether this arrangement extends to areas outside the frontal lobe, we examined subcortical inputs to a region of inferotemporal cortex, area TE (Middleton & Strick, 1996a) (figure 38.1, area TE). Area TE is known to project to the input stage of basal ganglia processing (i.e., the tail of the caudate and ventral portions of the putamen) (Saint-Cyr, Ungerleider, & Desimone, 1990) but not to the input stage of cerebellar processing (Glickstein et al., 1985; Schmahmann & Pandya, 1997).

Virus injections into area TE did not result in any labeled neurons in the deep cerebellar nuclei. This suggests that TE neither projects to nor receives from the cerebellum. On the other hand, the same virus injections into TE did result in a distinct cluster of labeled neurons in SNpr (figure 38.7, area TE). Most of these neurons were located dorsally in a caudal region of the SNpr that appears to be separate from the regions that influence the FEF or subdivisions of prefrontal cortex. Thus TE is both a source of input to and a target of output from a distinct portion of the basal ganglia.

TE is known to play a critical role in the visual recognition and discrimination of objects (e.g., Gross, 1972; Tanaka, Saito, Fukuda, & Moriya, 1991; Miyashita, 1993). Physiological studies have shown that the region of the SNpr that influences TE contains some neurons that are responsive to the presentation of visual stimuli (e.g., Hikosaka & Wurtz, 1983a). These observations, together with our anatomical results, provide evidence that basal ganglia output is involved in higher-order aspects of visual processing, as well as in motor and cognitive function.

Macroarchitecture of subcortical loops with the cerebral cortex

Our observation that cortical areas that receive output from the basal ganglia and cerebellum also project to the input stage of these subcortical structures suggests that closed loop circuits represent a fundamental architectural feature of basal ganglia and cerebellar connections with the cerebral cortex. We used retrograde transneuronal transport of rabies virus to define the region of cerebellar cortex that projects to a specific region of the cerebral cortex. Then we used anterograde transneuronal transport of the H129 strain of HSV1 to define the region of the cerebellar cortex that receives input from a specific region of the cerebral cortex (Kelly & Strick, 2003). Our first experiments using this approach examined the topographic organization of circuits linking the cerebellar cortex with the arm area of M1 and with area 46 in dorsolateral prefrontal cortex. In short, we found that the arm area of M1 receives input from Purkinje cells located mainly in lobules IV–VI of cerebellar cortex (figure 38.8, left). In contrast, area 46 receives input from Purkinje cells located mainly in crus II of the ansiform lobule (figure 38.8, right). Thus M1 and area 46 are the target of output from separate regions of the cerebellar cortex.

Anterograde transneuronal transport of the H129 strain HSV1 revealed that granule cells in lobules IV–VI of cerebellar cortex receive input from the arm area of M1. This is the same region of the cerebellar cortex that projects to M1 (figure 38.8, left). Similarly, granule cells in crus II of the cerebellar cortex receive input from area 46. This is the same region of the cerebellar cortex that projects to area 46 (figure 38.8, right). These observations provide strong support for our proposal that multiple closed loop circuits represent a fundamental architectural feature of cerebrocerebellar interactions. Similar closed loop circuits are also likely to be a fundamental feature of cerebrobasal ganglia interactions (Kelly & Strick, 2004).

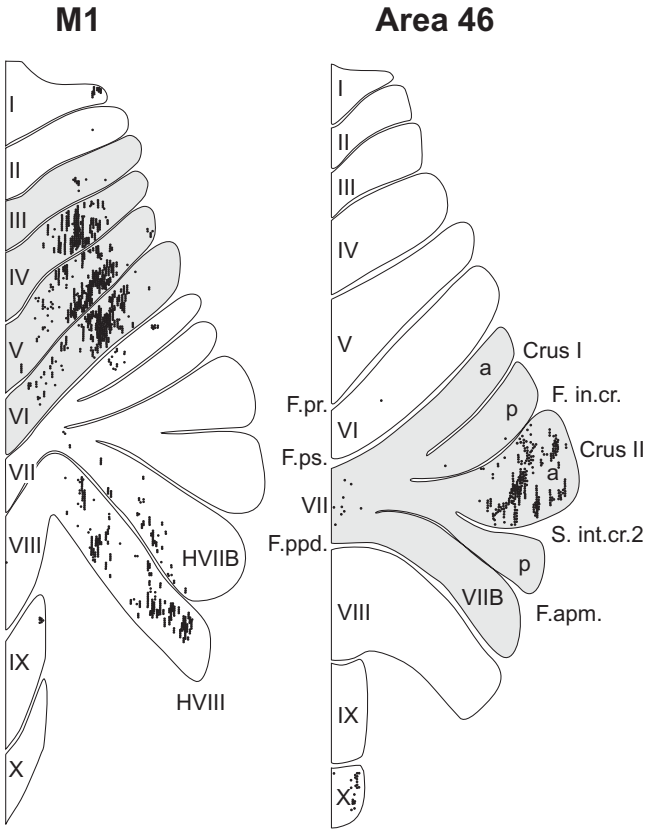


FIGURE 38.8 Origin of cerebellar cortical projections to M1 and area 46. The *black dots* on the flattened surface maps of the cerebellar cortex represent Purkinje cells that were labeled by retrograde transneuronal transport of rabies virus from the arm area of M1 (*left panel*) or from area 46 (*right panel*). Note that the Purkinje cells that project to M1 are located in separate lobules from those that project to area 46. Nomenclature and abbreviations are according to Larsell (1970). (Adapted from Kelly & Strick, 2003.)

Functional implications

Clearly, the outputs from the basal ganglia and cerebellum gain access to more widespread and diverse areas of cortex than was previously imagined. To date, our studies have shown that the output nuclei of the basal ganglia and cerebellum project (via the thalamus) to skeletomotor, oculomotor, prefrontal, and posterior parietal areas of cortex. In addition, a portion of SNpr projects to inferotemporal cortex. Thus, the anatomical substrate exists for the basal ganglia and cerebellum to influence higher-order aspects of cognition such as planning, working memory, sequential behavior, visuospatial perception, and attention as well as skeletomotor and oculomotor function. As a consequence, a sizable component of basal ganglia and cerebellar output operates outside of the domain of motor control.

Some support for this conclusion comes from recent analyses of the consequences of cerebellar pathology in human subjects. In addition to the classical motor deficits, there is considerable evidence that cerebellar damage can lead to deficits in the performance of cognitive tasks that require rule-based learning, judgment of temporal intervals, visuospatial analysis, shifting attention between sensory modalities, and working memory and planning (see reviews by Leiner et al., 1986, 1991, 1993; Botez, Botez, Elie, & Attig, 1989; Ivry & Keele, 1989; Fiez, Petersen, Cheney, & Raichle, 1992; Akshoomoff & Courchesne, 1992; Grafman et al., 1992; Schmahmann, 1991, 1997; Schmahmann & Sherman, 1998). Many of these deficits reflect functions that are normally thought to be subserved by areas of prefrontal cortex.

On the basis of our results, one interpretation of the origin of these deficits is that they result from an interruption of input to prefrontal cortex from the cerebellum. A study by Fiez and colleagues (1992) provides some support for this interpretation. They described a patient, designated RC1, who had circumscribed damage to the lateral portion of his right cerebellar cortex. This patient exhibited few classical signs of cerebellar damage but was impaired on the performance of specific types of rule-based language and memory tasks. The deficits appeared on tasks that in normal subjects activate lateral portions of the cerebellar hemispheres and areas 9 and 46 (Petersen, Fox, Posner, Mintun, & Raichle, 1988; Raichle et al., 1994; Fiez et al., 1996). Our anatomical studies suggest that the portions of the cerebellum that are damaged in RC1 are part of the cerebellar loop with the prefrontal cortex (Kelly & Strick, 2003). Thus the cognitive deficits in RC1 may have been a consequence of interrupting this circuit.

In general, we found that basal ganglia and cerebellar projections to a cortical area originate from a localized cluster of neurons that we have termed an output channel. The output channels to different cortical areas display a surprising degree of topographic organization. For example, the output channels that influence dorsomedial regions of prefrontal cortex are located largely in the GPi, whereas the output channels that influence ventrolateral regions of prefrontal cortex are located largely in the SNpr (Middleton & Strick, 2002; Akkal et al., 2007). Both sets of output channels are separate from those that influence skeletomotor and oculomotor areas of cortex. Output channels within the dentate are as topographically organized as those in the basal ganglia, if not more so (Dum & Strick, 2003; Akkal et al., 2007).

Evidence for a segregation of function in the human GPi comes from the observation that the cognitive and motor effects of pallidotomies, performed to ameliorate the symptoms of Parkinson's disease, depend significantly on the location of the lesion (Lombardi et al., 2000). Lesions located in the most anteromedial region of the GPi, the likely origin of output channels to prefrontal cortex, produced the greatest degree of cognitive impairment. In contrast, lesions in the intermediate region of the GPi, the likely origin of output channels to motor areas of cortex, led to maximal effects on motor performance but produced little effect on cognition. Thus the human GPi appears to have spatially separate motor and cognitive output channels.

To date, we have identified the output channels in the basal ganglia and cerebellum to skeletomotor, oculomotor, prefrontal, and some posterior parietal areas of cortex. All together, these output channels occupy approximately 70% of the volume of these subcortical nuclei. This means that the cortical targets for approximately 30% of the output from the basal ganglia and cerebellum remain to be identified. The architecture of basal ganglia and cerebellar loops with the cerebral cortex allows us to make some predictions about the identity of these targets. Cingulate, orbital frontal, and medial posterior parietal cortex are known to be major sources of input to the basal ganglia and cerebellum. Our results suggest that cortical areas that project to the input stage of the basal ganglia and cerebellum processing are the targets of the output stage of processing in these circuits. If this proposal is correct, then the remaining 30% of the basal ganglia and cerebellar output is directed at cingulate, orbital frontal, and medial posterior parietal areas of cortex. This prediction will be tested in future experiments.

The new insights gained from virus tracing have important implications for hypotheses about basal ganglia and cerebellar contributions to normal and abnormal behavior. Detailed discussions of this issue have been presented in our recent papers (Middleton & Strick, 1996a, 2001, 2002; Clower et al., 2001, 2005; Akkal et al., 2007); therefore only some examples will be presented here. It is known that abnormal activity in basal ganglia and cerebellar loops with motor areas of cortex results in striking disorders of movement. Likewise, abnormal activity in basal ganglia and cerebellar loops with nonmotor areas of the cerebral

cortex could lead to a broad range of psychiatric and neurological symptoms such as those associated with depression, obsessive-compulsive disorder, Parkinson's disease, and Huntington's disease (for a review, see Lichter & Cummings, 2000). For example, Courchesne and colleagues (1988, 1994, 1997) have suggested that alterations in the cerebellum and its projections to posterior parietal cortex may underlie some of the deficits seen in autistic patients. Rapoport and Wise (1988) have proposed that dysfunction in basal ganglia circuits with anterior cingulate and orbital frontal cortex may explain some of the features of obsessive-compulsive disorder. We have suggested that abnormal signals in the basal ganglia loop with area TE in inferotemporal cortex are responsible for the visual hallucinations that are seen in L-dopa toxicity (Middleton & Strick, 1996a). It is clear that additional studies aimed at unraveling these loops could lead to new insights into the pathophysiological basis of basal ganglia and cerebellar disorders.

In summary, virus tracing has revealed that the output of the basal ganglia and cerebellum targets motor, premotor, prefrontal, posterior parietal, and inferotemporal areas of cortex. These connections provide the basal ganglia and cerebellum with the anatomical substrate to influence not only the control of movement, but also many aspects of cognitive behavior such as planning, working memory, sequential behavior, visuospatial perception, and attention. Similarly, there is growing evidence that disorders such as schizophrenia, autism, attention-deficit disorder, and obsessive-compulsive disorder are associated with alterations in basal ganglia or cerebellar function. Thus, it is possible that abnormal activity in specific basal ganglia and cerebellar loops with the cerebral cortex results in identifiable sets of neuropsychiatric symptoms. Taken together, the recent findings about basal ganglia and cerebellar circuitry provide a new anatomical framework for understanding the contributions of these structures to diverse aspects of motor and nonmotor behavior.

ACKNOWLEDGMENTS This work was supported in part by funds from the Office of Research and Development, Medical Research Service, Department of Veterans Affairs and U.S. Public Health Service Grants R01 NS24328, R01 MH56661, P40 RR018604 (PLS).

REFERENCES

AKKAL, D., DUM, R. P., & STRICK, P. L. (2007). Supplementary motor area and presupplementary motor area: Targets of basal ganglia and cerebellar output. *J. Neurosci.*, *27*, 10659–10673.

AKSHOOMOFF, N. A., & COURCHESNE, E. (1992). A new role for the cerebellum in cognitive function. *Behav. Neurosci.*, *106*, 731–738.

ALEXANDER, G. E., DELONG, M. R., & STRICK, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.*, *9*, 357–381.

ALLEN, G. I., & TSUKAHARA, N. (1974). Cerebrocerebellar communication systems. *Physiol. Rev.*, *54*, 957–1006.

BHATIA, K. P., & MARSDEN, C. D. (1994). The behavioural and motor consequences of focal lesions of the basal ganglia in man. *Brain*, *117*, 859–876.

BOTEZ, M. I., BOTEZ, T., ELIE, R., & ATTIG, E. (1989). Role of the cerebellum in complex human behavior. *Ital. J. Neurol. Sci.*, *10*, 291–300.

BROOKS, V. B., & THACH, W. T. (1981). Cerebellar control of posture and movement. In V. B. Brooks (Ed.), *Handbook of physiology: Section 1: The nervous system, Vol. 2: Motor control, Part II* (pp. 877–946). Bethesda, MD: American Physiological Society.

CAVADA, C., & GOLDMAN-RAKIC, P. S. (1991). Topographic segregation of corticostriatal projections from posterior parietal subdivisions in the macaque monkey. *Neuroscience*, *42*, 683–696.

CLOWER, D. M., DUM, R. P., & STRICK, P. L. (2005). Basal ganglia and cerebellar inputs to "AIP." *Cereb. Cortex*, *15*, 913–920.

CLOWER, D. M., WEST, R. A., LYNCH, J. C., & STRICK, P. L. (2001). The inferior parietal lobule is the target of output from the superior colliculus, hippocampus and cerebellum. *J. Neurosci.*, *21*, 6283–6291.

COURCHESNE, E. (1997). Brainstem, cerebellar and limbic neuroanatomical abnormalities in autism. *Curr. Opin. Neurobiol.*, *7*, 269–278.

COURCHESNE, E., TOWNSEND, J., AKSHOOMOFF, N., SAITOH, O., YEUNG-COURCHESNE, R., LINCOLN, A., JAMES, H., HAAS, R., SCHREIBMAN, L., & LAU, L. (1994). Impairment in shifting attention in autistic and cerebellar patients. *Behav. Neurosci.*, *108*, 848–865.

COURCHESNE, E., YEUNG-COURCHESNE, R., PRESS, G. A., HESSELINK, J. R., & JERNIGAN, T. L. (1988). Hypoplasia of cerebellar vermal lobules VI and VII in autism. *N. Engl. J. Med.*, *318*, 1349–1354.

DELONG, M. R., & GEORGOPOULOS, A. P. (1981). Motor functions of the basal ganglia. In V. B. Brooks (Ed.), *Handbook of physiology: Section I: The nervous system, Vol. II: Motor control* (pp. 1017–1061). Bethesda, MD: American Physiological Society.

DUM, R. P., & STRICK, P. L. (2003). An unfolded map of the cerebellar dentate nucleus and its projections to the cerebral cortex. *J. Neurophysiol.*, *89*, 634–639.

FIEZ, J. A., PETERSEN, S. E., CHENEY, M. K., & RAICHLE, M. E. (1992). Impaired non-motor learning and error detection associated with cerebellar damage. *Brain*, *115*, 155–178.

FIEZ, J. A., RAIFE, E. A., BALOTA, D. A., SCHWARZ, J. P., RAICHLE, M. E., & PETERSEN, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *J. Neurosci.*, *16*, 808–822.

FUSTER, J. M. (1997). *The prefrontal cortex*. New York: Raven Press.

GLICKSTEIN, M., MAY, J. G., & MERCIER, B. E. (1985). Corticopontine projection in the macaque: The distribution of labelled cortical cells after large injections of horseradish peroxidase in the pontine nuclei. *J. Comp. Neurol.*, *235*, 343–359.

GOLDMAN-RAKIC, P. S. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *351*, 1445–1453.

GOLDMAN-RAKIC, P. S., & SELEMON, L. D. (1990). New frontiers in basal ganglia research: Introduction. *Trends Neurosci.*, *13*, 241–244.

GRAFMAN, J., LITVAN, I., MASSAQUOI, S., STEWART, M., SIRIGU, A., & HALLETT, M. (1992). Cognitive planning deficit in patients with cerebellar atrophy. *Neurology*, *42*, 1493–1496.

Gross, C. G. (1972). Visual functions of inferotemporal cortex. In R. Jung (Ed.), *Handbook of sensory physiology* (pp. 451–482). Berlin: Springer-Verlag.

Hikosaka, O., & Wurtz, R. H. (1983a). Visual and oculomotor functions of monkey substantia nigra pars reticulata: I. Relation of visual and auditory responses to saccades. *J. Neurophysiol.*, *49*, 1230–1253.

Hikosaka, O., & Wurtz, R. H. (1983b). Visual and oculomotor functions of monkey substantia nigra pars reticulata: III. Memory-contingent visual and saccade responses. *J. Neurophysiol.*, *49*, 1268–1284.

Hoover, J. E., & Strick, P. L. (1993). Multiple output channels in the basal ganglia. *Science*, *259*, 819–821.

Hoover, J. E., & Strick, P. L. (1999). The organization of cerebello- and pallido-thalamic projections to primary motor cortex: An investigation employing retrograde transneuronal transport of herpes simplex virus type 1. *J. Neurosci.*, *19*, 1446–1463.

Inase, M., & Tanji, J. (1995). Thalamic distribution of projection neurons to the primary motor cortex relative to afferent terminal fields from the globus pallidus in the macaque monkey. *J. Comp. Neurol.*, *353*, 415–426.

Ivry, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *J. Cogn. Neurosci.*, *1*, 136–152.

Jinnai, K., Nambu, A., Tanibuch, I., & Yoshida, S. (1993). Cerebello- and pallido-thalamic pathways to areas 6 and 4 in the monkey. *Stereotactic Funct. Neurosurg.*, *60*, 70–79.

Karnath, H. O., Himmelbach, M., & Rorden, C. (2002). The subcortical anatomy of human spatial neglect: Putamen, caudate nucleus and pulvinar. *Brain*, *125*, 350–360.

Kelly, R. M., & Strick, P. L. (2000). Rabies as a transneuronal tracer of circuits in the central nervous system. *J. Neurosci. Methods*, *103*, 63–71.

Kelly, R. M., & Strick, P. L. (2003). Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *J. Neurosci.*, *12*, 8432–8444.

Kelly, R. M., & Strick, P. L. (2004). Macro-architecture of basal ganglia loops with the cerebral cortex: Use of rabies virus to reveal multisynaptic circuits. *Progr. Brain Res.*, *143*, 449–459.

Kemp, J. M., & Powell, T. P. S. (1971). The connexions of the striatum and globus pallidus: Synthesis and speculation. *Phil. Trans. R. Soc. Lond. B Biol. Sci.*, *262*, 441–457.

Larsell, O. (1970). *The comparative anatomy and histology of the cerebellum from monotremes through apes*. Minneapolis: University of Minnesota Press.

Leiner, H. C., Leiner, A. L., & Dow, R. S. (1986). Does the cerebellum contribute to mental skills? *Behav. Neurosci.*, *100*, 443–454.

Leiner, H. C., Leiner, A. L., & Dow, R. S. (1991). The human cerebro-cerebellar system: Its computing, cognitive, and language skills. *Behav. Brain Res.*, *44*, 113–128.

Leiner, H. C., Leiner, A. L., & Dow, R. S. (1993). Cognitive and language functions of the human cerebellum. *Trends Neurosci.*, *16*, 444–447.

Lichter, D. G., & Cummings, J. L. (2000). *Frontal-subcortical circuits in psychiatry and neurology*. New York: Guilford.

Lombardi, W. J., Gross, R. E., Trepanier, L. L., Lang, A. E., Lozano, A. M., & Saint-Cyr, J. A. (2000). Relationship of lesion location to cognitive outcome following microelectrode-guided pallidotomy for Parkinson's disease: Support for the existence of cognitive circuits in the human pallidum. *Brain*, *123*, 746–758.

Lynch, J. C., Hoover, J. E., & Strick, P. L. (1994). Input to the primate frontal eye field from the substantia nigra, superior colliculus, and dentate nucleus demonstrated by transneuronal transport. *Exp. Brain Res.*, *100*, 181–186.

Middleton, F. A., & Strick, P. L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, *266*, 458–461.

Middleton, F. A., & Strick, P. L. (1996a). The temporal lobe is a target of output from the basal ganglia. *Proc. Natl. Acad. Sci. USA*, *93*, 8683–8687.

Middleton, F. A., & Strick, P. L. (1996b). New concepts regarding the organization of basal ganglia and cerebellar ouput. In M. Ito & Y. Miyashita (Eds.), *Integrative and molecular approach to brain function* (pp. 253–271). New York: Elsevier Science.

Middleton, F. A., & Strick, P. L. (1997). Cerebellar output channels: Substrates for the control of motor and cognitive function. In J. Schmahmann (Ed), *The cerebellum and cognition* (vol. 41, pp. 61–82.) San Diego: Academic.

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia and cerebellar loops: Motor and cognitive circuits. *Brain Res. Brain Res. Rev.*, *31*, 236–250.

Middleton, F. A., & Strick, P. L. (2001). Cerebellar projections to the prefrontal cortex of the primate. *J. Neurosci.*, *21*, 700–712.

Middleton, F. A., & Strick, P. L. (2002). Basal ganglia "projections" to the prefrontal cortex. *Cereb. Cortex*, *12*, 926–935.

Miyashita, Y. (1993). Inferior temporal cortex: Where visual perception meets memory. *Annu. Rev. Neurosci.*, *16*, 245–263.

Mushiake, H., & Strick, P. L. (1993). Preferential activity of dentate neurons during limb movements. *J. Neurophysiol.*, *70*, 2660–2664.

Mushiake, H., & Strick, P. L. (1995). Pallidal neuron activity during sequential arm movements. *J. Neurophysiol.*, *74*, 2754–2758.

Passingham, R. (1993). The frontal lobes and voluntary action. Oxford, UK: Oxford University Press.

Percheron, G., François, C., Talbi, B., Yelnik, J., & Fénelon, G. (1996). The primate motor thalamus. *Brain Res. Rev.*, *22*, 93–181.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*, 585–589.

Picard, N., & Strick, P. L. (2001). Imaging the premotor areas. *Curr. Opin. Neurobiol.*, *11*, 663–672.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. M., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb. Cortex*, *4*, 8–26.

Rapoport, J. L., & Wise, S. P. (1988). Obsessive-compulsive disorder: Evidence for basal ganglia dysfunction. *Psychopharm. Bull.*, *24*, 380–384.

Saint-Cyr, J. A., Ungerleider, L. G., & Desimone, R. (1990). Organization of visual cortical inputs to the striatum and subsequent outputs to the pallido-nigral complex in the monkey. *J. Comp. Neurol.*, *298*, 129–156.

Sakai, S. T., Inase, M., & Tanji, J. (1999). Pallidal and cerebellar inputs to thalamocortical neurons projecting to the supplementary motor area in Macaca fuscata: A triple-labeling light microscopic study. *Anat. Embryol. (Berl.)*., *199*, 9–19.

Schell, G. R., & Strick, P. L. (1984). The origin of thalamic inputs to the arcuate premotor and supplementary motor areas. *J. Neurosci.*, *4*, 539–560.

Schmahmann, J. D. (1991). An emerging concept: The cerebellar contribution to higher function. *Arch. Neurol.*, *48*, 1178–1187.

Schmahmann, J. D. (1997). Rediscovery of an early concept. *Int. Rev. Neurobiol.*, *41*, 3–27.

Schmahmann, J. D., & Pandya, D. N. (1997). The cerebrocerebellar system. *Int. Rev. Neurobiol.*, *41*, 31–60.

Schmahmann, J. D., & Sherman, J. C. (1998). The cerebellar cognitive affective syndrome. *Brain*, *121*, 561–579.

Strick, P. L., & Card, J. P. (1992). Transneuronal mapping of neural circuits with alpha herpesviruses. In J. P. Bolam (Ed.), *Experimental neuroanatomy: A practical approach* (pp. 81–101). Oxford, UK: Oxford University Press.

Strick, P. L., Dum, R. P., & Picard, N. (1995). Macro-organization of the circuits connecting the basal ganglia with the cortical motor areas. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 117–130). Cambridge, MA: MIT Press.

Tanaka, K., Saito, H.-A., Fukuda, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey, *J. Neurophysiol.*, *66*, 170–189.

van Kan, P. L. E., Houk, J. C., & Gibson, A. R. (1993). Output organization of intermediate cerebellum of the monkey. *J. Neurophysiol.*, *69*, 57–73.

Yeterian, E. H., & Pandya, D. N. (1993). Striatal connections of the parietal association cortices in rhesus monkeys. *J. Comp. Neurol.*, *332*, 175–197.

39 The Basal Ganglia and Cognition

ANN M. GRAYBIEL AND JONATHAN W. MINK

ABSTRACT Clinical evidence, experimental studies in animals, and anatomical findings suggest that the basal ganglia act to influence not only motor behavior but also cognitive functions. We discuss the functions of the basal ganglia in relation to four categories: (1) movement release and inhibition, (2) response selection, (3) attention and assignment of salience, and (4) learning and adaptive control of behavior. In establishing these functions, striatal output neurons lead into different output pathways: the direct, indirect, hyperdirect, and striosomal pathways. Divergence of cortical inputs to the striatum and reconvergence of these motor and cognitive signals in cortico-basal ganglia pathways is seen as essential in remapping forebrain representations of action and intrastriatal networks in the binding process. We propose that a crucial feature of this remapping is a learning-related recoding of sequential motor and cognitive action representations so that they can be expressed as units. This chunking function of the striatum and associated cortico-basal ganglia loops may be a key mechanism operative across each of the functional categories of behavioral control attributed to the basal ganglia.

The basal ganglia make up a group of interconnected subcortical nuclei that are organized into circuits involved in the control of behavior. The basal ganglia have long been recognized as important for motor control, because prominent movement disorders such as Parkinson's disease result from basal ganglia dysfunction. However, it is now widely recognized that the basal ganglia are parts of cortico-thalamo-basal ganglia loops, and that they function not only in sensorimotor control, but also in a wide range of cognitive processes ranging from attention to emotion, from response release and inhibition to response selection, and from on-line control to a primary function in learning and memory. Accordingly, the basal ganglia have now been implicated in an equally broad range of clinical disorders, ranging from the classical extrapyramidal disorders (Parkinson's disease, Huntington's disease, and dystonia) to neuropsychiatric disorders including obsessive-compulsive disorder, Tourette syndrome, attention-deficit disorder, and even schizophrenia. The basal ganglia themselves contain highly organized neuronal circuitry, but they do not function in isolation. The basal ganglia are intimately connected with the cerebral cortex and are perhaps best viewed as parts of cortico-basal ganglia circuits.

How could this system have such broad functions? A clue that the basal ganglia might contribute to cognitive processing is that the basal ganglia attain a very large size in the human brain. But an even more telling clue is that a large part of the outflow of the basal ganglia in primates is directed via the thalamus toward executive areas of the frontal cortex—areas that are themselves associated with attention, planning, volitional decision, and selection among potential responses to external or internal cues (Fuster, 1997; Paus, 2001). Yet more evidence comes from brain imaging studies of subjects engaged in cognitive tasks (Klein, Zatorre, Milner, Meyer, & Evans, 1994; Grafton, Hazeltine, & Ivry, 1995; Braver et al., 1997; Rao et al., 1997; Desmond, Gabrieli, & Glover, 1998; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Peigneux et al., 2000; Poldrack & Gabrieli, 2001; Small, Zatorre, Daghler, Evans, & Jones-Gotman, 2001; van den Heuvel et al., 2003; Cools, Ivry, & Esposito, 2006; Chang, Crottaz-Herbette, & Menon, 2007; Cools, Gibbs, Miyakawa, Jagust, & D'Esposito, 2008; Dahlin, Neely, Larson, Backman, & Nyberg, 2008; McNab & Klingberg, 2008) and from findings in patients with brain dysfunction due to disease or injury (Mendez, Adams, & Lewandowski, 1989; Bhatia & Marsden, 1994; Sawamoto et al., 2007). Experiments on animals have also generated working hypotheses about the neurobiology underlying basal ganglia function (Oberg & Divac, 1979; Graybiel, 1995, 1998, 2005, 2008; Miyashita, Hikosa, & Kato, 1995; Bergman et al., 1998; Hikosaka et al., 1999; Jog, Kubota, Connolly, Hillegart, & Graybiel, 1999; Brainard & Doupe, 2000; Mink, 2001; Packard & Knowlton, 2002; Barnes, Kubota, Hu, Jin, & Graybiel, 2005; Apicella, 2007). Together, these findings have brought the basal ganglia to the forefront of work on how the brain engages in interactions with the sensory and internal environment to form structured predictions about the world and, on this basis, to make and execute action plans (figures 39.1 and 39.2).

ANN M. GRAYBIEL Department of Brain and Cognitive Sciences and the McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

JONATHAN W. MINK Departments of Neurology, Neurobiology and Anatomy, Brain and Cognitive Sciences, and Pediatrics, University of Rochester School of Medicine and Dentistry, Rochester, New York

Perspectives from anatomy

The basal ganglia receive a massive input from the neocortex. Most of these cortical inputs are directed toward

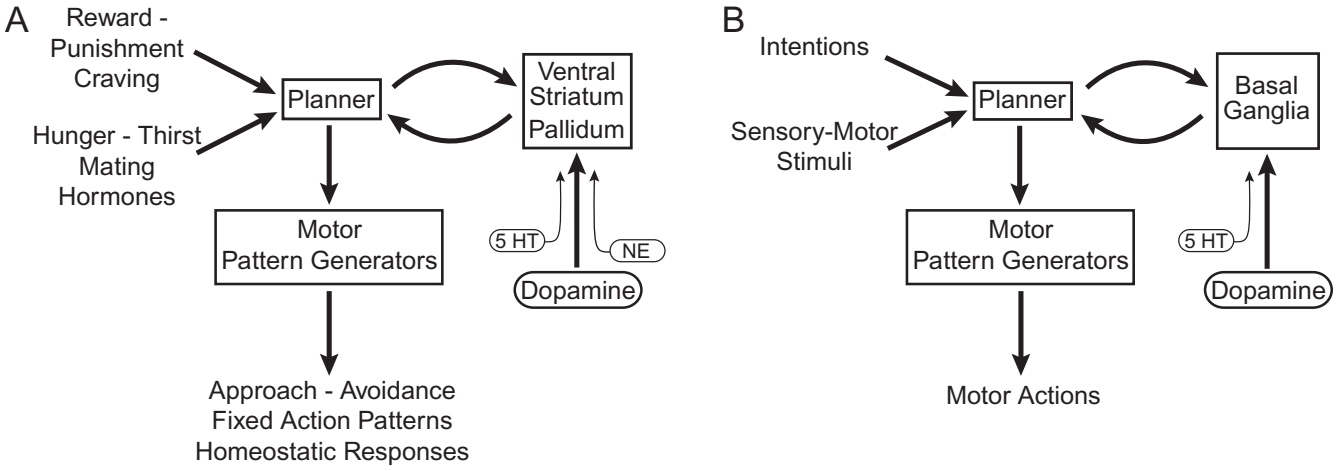


FIGURE 39.1 Diagrams illustrating the postulated functions of the basal ganglia in relation to central pattern generators for eliciting goal-directed behavior (*A*) and movement (*B*). Planner circuits of the forebrain are influenced by motivation-related inputs (*A*) and sensory-motor stimuli (*B*). Abbreviations: 5-HT, serotonin; NE, norepinephrine. (Adapted from Graybiel, 1997.)

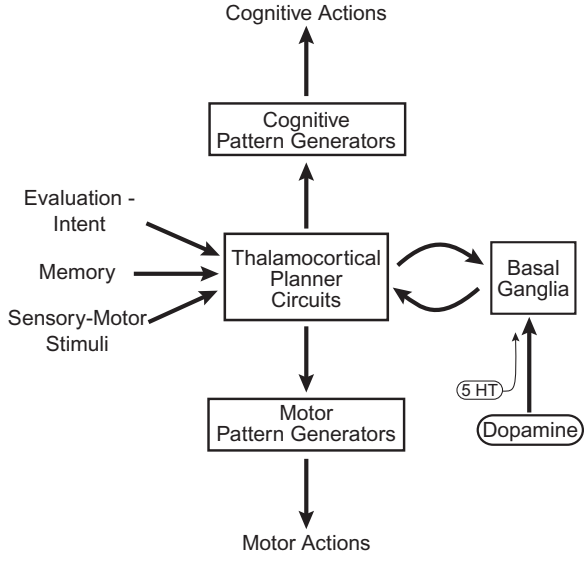


FIGURE 39.2 Schematic diagram illustrating potential influences of the basal ganglia not only on motor pattern generators but also on cognitive pattern generators. (Adapted from Graybiel, 1997.)

the striatum (caudate nucleus and putamen). The range of these corticostriatal inputs is impressive. They come not only from primary and higher-order sensory areas and from motor and premotor areas, but also from the large areas of association cortex in the parietal, temporal, medial, and frontal association cortex (Webster, Bachevalier, & Ungerleider, 1993; Eblen & Graybiel, 1995; Yeterian & Pandya, 1998; Ferry, Ongur, An, & Price, 2000; Leichnetz, 2001; Haber, Kim, Mailly, & Calzavara, 2006; Calzavara, Mailly, & Haber, 2007). There are other very large inputs to the striatum from thalamic nuclei, especially the intralaminar nuclei (Parent, Mackey, & De Bellefeuille, 1983; Ragsdale & Graybiel, 1991; Haber & McFarland, 2001). Further inputs come to other nuclei in basal ganglia circuits, as we will see below. If one includes, as should be done, the ventral striatum/ventral pallidum in the basal ganglia, cortical inputs to the system also come from the hippocampal formation and amygdala (Groenewegen, Wright, & Uylings, 1997; Fudge, Kunishio, Walsh, Richard, & Haber, 2002). When we add inputs from neuromodulatory systems, including the dopamine-containing nigrostriatal tract and serotonergic inputs, and inputs from the neocortex and elsewhere to other nuclei of the basal ganglia, the inputs to the basal ganglia system as a whole are rich and diverse and by no means restricted to one functional domain.

It is also important to keep in mind that different regions within each nucleus of the basal ganglia are probably as different from one another functionally as different parts of the neocortex are from one another. When we think of behavior-related functions of the neocortex, we naturally think of the functions of individual cortical areas, for example, the middle temporal area for visual motion, parietal areas for reach and grasp, or prefrontal areas for working memory. We do not know as much about the regionally specialized subdivisions of the basal ganglia, but it is clear that there are different "families" of cortico-basal ganglia circuits related to motor, associative, and limbic functions (Graybiel, 1984; Alexander, DeLong, & Strick, 1986). The behavioral evidence leading to this idea is important: Lesion studies have shown that localized lesions of the striatum produce symptoms similar to those induced by lesions in the cortical areas projecting strongly to the particular parts of the damaged striatum (Divac, Rosvold, & Szwarcbart, 1967; Goldman & Rosvold, 1972). Somehow, the functional domain of specific basal ganglia circuits seems to relate to the function of their cortical input sources.

What the basal ganglia do, of course, depends not only on their inputs but also on their outputs. Here, the story is interesting. Most current anatomical tract-tracing studies indicate that the largest ascending outflow of the basal ganglia is directed toward the frontal cortex via synaptic links in the thalamus. This puts the basal ganglia squarely in the "executive" realm of function. This view is consistent with the undoubted participation of these structures in motor control. But the frontal areas that receive basal ganglia outflow extend from the classical motor and premotor areas into the prefrontal cortex (Middleton & Strick, 2002). In fact, a large part of the neocortex in front of the central sulcus—including medial and lateral prefrontal, cingulate, and lateral and orbitofrontal cortex—is now thought to receive inputs (via thalamically processed routes) from the basal ganglia proper. Thus the outflow from the basal ganglia reaches regions that function in cognitive and emotional control. Anatomical work by Strick and colleagues with viral transport methods demonstrates that also part of the inferotemporal cortex and part of the parietal cortex also receive inputs via basal ganglia–thalamocortical pathways (see chapter 39). These temporal and parietal areas are themselves linked to the executive and cognitive networks of the frontal lobes. Conceptually, then, the anatomy overwhelmingly favors the basal ganglia as poised to influence high-level functions associated with activity in the frontal neocortex (figure 39.3).

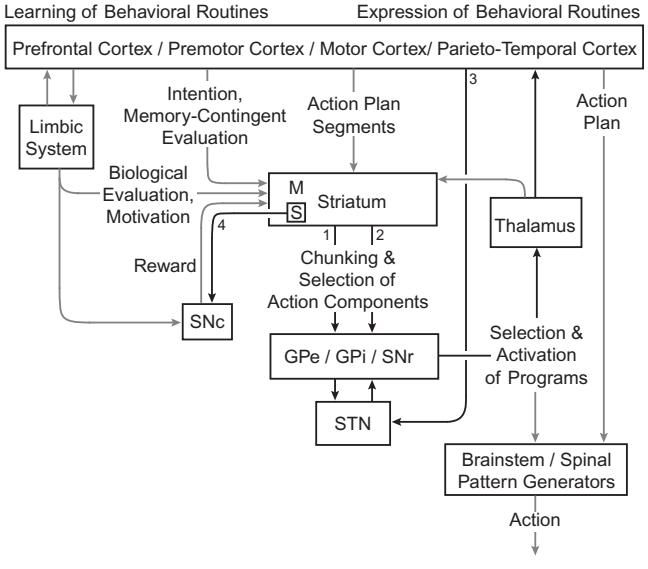


FIGURE 39.3 Schematic diagram of major basal ganglia circuits with highly schematized indications of component functions. The striatum with its matrix (M) and striosomal (S) compartments is centered in the diagram. Four major pathways are emphasized: the direct (1) and indirect (2) pathways, the hyperdirect pathway (3), and the striosomal pathway (4). Abbreviations: GPe, GPi, globus pallidus external and internal subdivisions; STN, subthalamic nucleus; SNc, dopamine-containing substantia nigra pars compacta. (Adapted from Graybiel, 1997.)

This emphasis on the cortically directed pathways leading out from the basal ganglia is natural enough when thinking about the possible cognitive functions of the basal ganglia, but it is equally important to keep in mind that there are other robust outputs of the basal ganglia (Graybiel & Ragsdale, 1979; Parent & Hazrati, 1995a, 1995b). These may also, directly or indirectly, influence potential cognitive functions of the basal ganglia. Such connections include projections to the reticular nucleus of the thalamus, a major controller of thalamocortical and corticothalamic state-dependent activity (McAlonan & Brown, 2002), and descending projections from the basal ganglia, among which are connections leading to the superior colliculus and brain stem reticular formation, and to the nuclei that are recurrently connected with the basal ganglia (Parent & Hazrati, 1995a, 1995b). The largest of these are the descending projections to the superior colliculus (Rinvik, Grofova, & Ottersen, 1976; Graybiel, 1978) and to the pedunculopontine area of the brain stem reticular formation (Parent & Hazrati, 1995a, 1995b).

The most intensively studied of the recurrent-pathway nuclei is the substantia nigra, and this is for a good reason. The dopamine-containing subdivision of the substantia nigra degenerates in Parkinson's disease and in related parkinsonian disorders. It is now known that dopamine-containing neurons respond phasically to predictors of reward or to primary rewards themselves and have tonic activity that appears to reflect probability of reward (Schultz, 2002; Fiorillo, Tobler, & Schultz, 2003). A second key nucleus is the subthalamic nucleus. The subthalamic nucleus, like the striatum, receives input from cerebral cortex, and as we will see, this nucleus is a pivotal control nucleus of the basal ganglia circuitry. It functions both as a direct cortical input node and as part of recurrent loops within basal ganglia circuits. Cortical inputs arise largely from the frontal lobes (Monakow, Akert, & Kunzle, 1978; Nambu, Yoshida, & Jinnai, 1990; Nambu, Takada, Inase, & Tokuno, 1996; Kolomiets et al., 2001). Lesions of this nucleus result in the hyperkinetic syndrome called ballism. Remarkably, it is now recognized that lesions or deep-brain stimulation in the subthalamic nucleus can relieve symptoms of Parkinson's disease (Bergman, Wichman, & DeLong, 1990; Lang, 2000; Obeso et al., 2000; Benabid et al., 2003). The input from the motor cortex (and related cortical areas) to the subthalamic nucleus is now recognized as the hyperdirect pathway (Nambu, Tokuno, & Takada, 2002), and this pathway is now considered to be key to the control of the motor functions—and also probably the cognitive and emotion-related functions—of the basal ganglia (Feger, Bevan, & Crossman, 1994; Deschenes, Bourassa, Doan, & Parent, 1996; Nambu et al., 2002; Schupbach & Agid, 2008) pathway. The third nucleus that we consider is the pedunculopontine nucleus, embedded in the reticular formation. The pedunculopontine

nucleus sends outputs to motor control centers of the lower brain stem and also is modulator of basal ganglia function by way of its recurrent upstream connections (Lavoie & Parent, 1994; Pahapill & Lozano, 2000). This nucleus is now also considered to be a key controller of basal ganglia output functions. Put together, then, we can see that the basal ganglia can influence a wide range of behavior through both their ascending projections toward the neocortex and by way of key descending connections to an influential set of brain stem nuclei.

Perspectives from the clinic

In Parkinson's disease and Huntington's disease, classical basal ganglia disorders that are accompanied by motor abnormalities, patients frequently experience cognitive dysfunction, including memory deficiency, depression, and, most important, disordered executive functions (R. G. Brown & Marsden, 1990; Dubois, Pillon, & Agid, 1992; Bédard et al., 1998, 2003; Joel, 2001; Saint-Cyr, 2003). Moreover, dementia is frequent in these diseases and mostly exhibits the features of frontal dementia syndromes (Pillon et al., 1994; Saint-Cyr, 2003). These "additional" (i.e., non-motor) symptoms are often taken to suggest that the basal ganglia proper have cognitive functions, but this view is now being modified by the realization that both Parkinson's disease and Huntington's disease are neurodegenerative diseases with neuronal damage extending—and even starting—outside the basal ganglia (Sieradzan & Mann, 2001; Braak et al., 2003). It is therefore not possible to attribute cognitive signs and symptoms of these disorders to dysfunction of the basal ganglia alone. However, there are other reasons to place the basal ganglia firmly within the cognitive domain.

First, there are other neurological disorders involving cognitive dysfunction that at least appear to affect more selectively specific basal ganglia nuclei (e.g., certain cerebrovascular disorders). Dysfunction in multiple cognitive domains is reported with infarctions or hemorrhages that are apparently limited to the caudate nucleus (e.g., abulia, restlessness, disinhibition and impulsivity, executive dysfunction) or to the putamen (e.g., contralateral neglect, language abnormalities) (Mendez et al., 1989; Caplan et al., 1990; Bhatia & Marsden, 1994). The single most prominent symptom associated with lesions of the caudate nucleus is abulia, a lack of drive. Lesions in the ventromedial caudate nucleus are associated with disinhibition and impulsivity. Work in the behaving primates supports such clinical findings and suggests some topographic differences of behavioral mechanisms within the basal ganglia (Francois et al., 2002; Tremblay et al., 2003).

Finally, imaging studies have demonstrated alterations of basal ganglia activity in disorders in which cognitive deficits are evident, including obsessive-compulsive (OC) spectrum disorders such as Tourette syndrome and obsessive-compulsive disorder (OCD), and attention-deficit/hyperactivity disorder (ADHD) (Graybiel & Rauch, 2000; Teicher et al., 2000; Rauch et al., 2001; Leckman, 2002; Albin & Mink, 2006). We will refer to these disorders again in what follows.

Hypotheses of basal ganglia function

We have grouped hypotheses about the functions of the basal ganglia into categories related to (1) movement release and inhibition, (2) response selection, (3) attention and assignment of salience, and (4) learning and adaptive control of behavior.

Movement Release and Inhibition We will in this chapter use the "motor circuits" of the basal ganglia as templates to understand the organization of basal ganglia circuits. Increasing evidence suggests that this same organization underlies the cognitive functions of the basal ganglia.

The main pathways that lead out from the basal ganglia are known as the direct and indirect pathways, the hyperdirect pathway, and the striosomal pathway (figures 39.3 and 39.4). These pathways are thought to influence motor control by release (direct pathway) or inhibition (indirect pathway and hyperdirect pathway) of motor behaviors (Albin, Young, & Penney, 1989; DeLong, 1990) and by control of the repetitiveness of behaviors for the striosomal pathway (Graybiel, Canales, & Capper-Loup, 2000; Saka & Graybiel, 2004). The release-inhibit model suggests, in simplest form, that the neocortex excites the striatum, which inhibits the internal pallidum, which in turn inhibits the motor thalamus. The double inhibition suggests that cortical activation phasically "releases" the thalamus, which then can excite the neocortex. This movement-releasing pathway is, according to the release-inhibit model, in direct competition with the *indirect pathway*, which, due to its connecting link in the subthalamic nucleus, is thought to depress movement. The subthalamic nucleus is released by striatal excitation of its inhibitory input nucleus (the external pallidum), and the subthalamic nucleus then excites the internal pallidum, leading to less movement. Many current models of the basal ganglia focus on this winner-take-all model (Dominey, Arbib, & Joseph, 1995; Beiser & Houk, 1998; J. E. Brown, Bullock, & Grossberg, 1999; Gillies & Arbuthnott, 2000; Frank, Loughry, & O'Reilly, 2001a; Kitano, Aoyagi, & Fukai, 2001). Current evidence suggests that this release-inhibit model may need revision (see Graybiel, 2005), but it has spurred major clinical advances and has been applied to the cognitive side of basal ganglia function as well.

The *hyperdirect pathway* (figure 39.3) consists of a fast, direct excitatory pathway from the motor cortex (and some other

cortical areas) to the subthalamic nucleus, which can therefore rapidly excite the pallidum (Nambu et al., 2002). This pathway bypasses the striatum and could help to account for the efficacy of deep brain stimulation of the subthalamic nucleus to relieve symptoms of Parkinson's disease (Lang, 2000; Obeso et al., 2000; Benabid et al., 2003).

We emphasize that if we extend the release-inhibit idea to the cognitive level, we can think of other actions—even thoughts or emotions—as being released through this mechanism (Swerdlow & Koob, 1987; Graybiel, 1997). This possibility is receiving potential support from the results of deep brain stimulation used as a therapeutic intervention for Parkinson's disease (Bejjani et al., 1999; Krack et al., 2001).

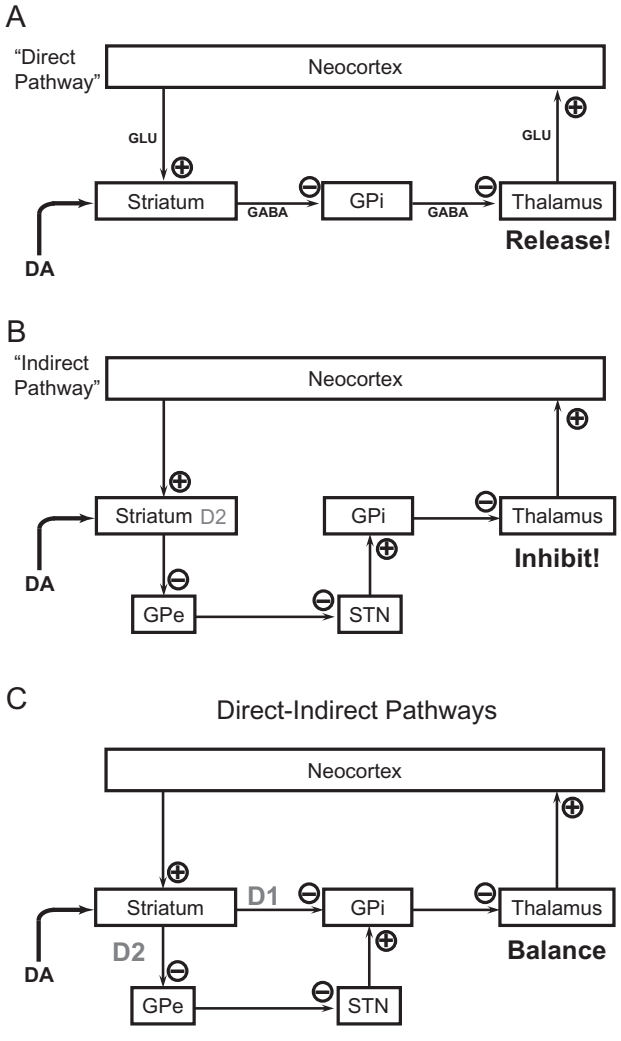


Figure 39.4 Highly schematic diagrams of the direct and indirect pathways identified in basal ganglia circuitry. *A* and *B* separate out these two pathways to emphasize their proposed "release" and "inhibit" functions. The diagrams in *C* puts them together to show the balance between them that is thought to underlie normal behavioral control. The hyperdirect and striosomal pathways are not shown here (see figures 39.3, 39.5, and 39.6).

The *striosomal pathway* (figure 39.3) is thought to lead from the anterior cingulate cortex and caudal orbitofrontal cortex to neurochemically defined compartments in the striatum that are called striosomes (striatal bodies). These in turn are interconnected with the dopamine-containing substantia nigra (Eblen & Graybiel, 1995; Prensa & Parent, 2001). These connections may serve to regulate the functions of the substantia nigra, and this striosomal system has been implicated regulating the frequency and repetitiveness of behavior of actions (Graybiel et al., 2000; Graybiel & Canales, 2001).

Selection and Inhibition of Competing Behaviors If we now think of these control pathways again, we can note that there are two primary disynaptic pathways of information flow from the cerebral cortex to the basal ganglia output nuclei, the internal pallidum (GPi), and the substantia nigra, pars reticulata (SNpr): (1) the disynaptic direct pathway from the neocortex to striatum to the output nuclei and (2) the hyperdirect pathway from the neocortex to the subthalamic nucleus (STN) and then to the basal ganglia output nuclei (figures 39.5 and 39.6). The contrasts between these two are important. First, the cortical input to the STN comes only from the frontal cortex, whereas the input to the striatum arises from all or nearly all areas of the cerebral cortex.

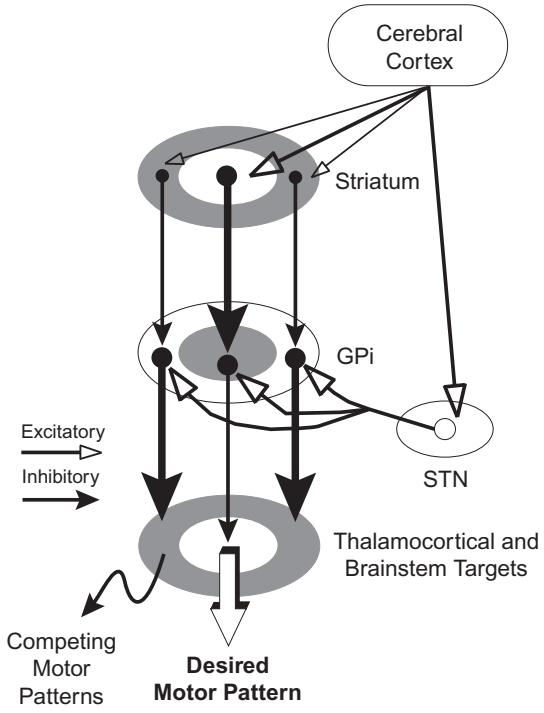


Figure 39.5 Schematic diagram of functional organization of the basal ganglia output. Excitatory projections are indicated with open arrows; inhibitory projections are indicated with filled arrows. Relative magnitude of activity is represented by line thickness. (Modified from Mink, 2001.)

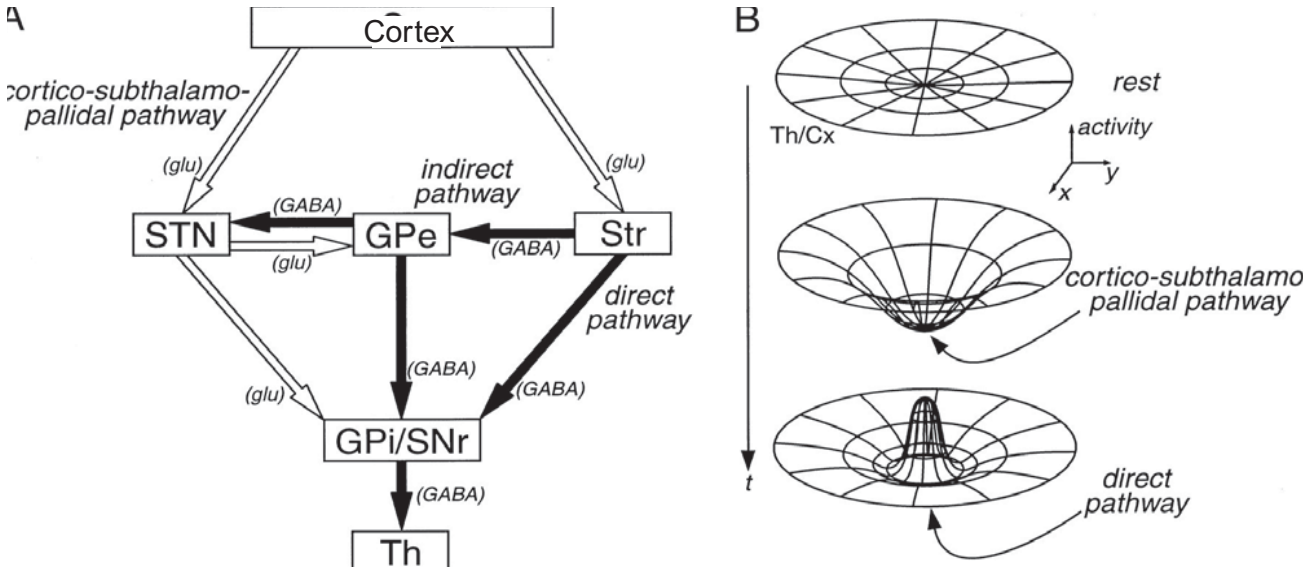


FIGURE 39.6 (*A*) Schematic diagram of the hyperdirect cortico-subthalamo-pallidal, the direct cortico-striato-pallidal, and the indirect corticostriato-GPe-subthalamo-GPi pathways. *White and black arrows* represent excitatory glutamatergic (glu) and inhibitory GABAergic (GABA) projections, respectively. Abbreviations: GPe, external segment of the globus pallidus; GPi, internal segment of the globus pallidus; SNr, substantia nigra, pars reticulata; STN, subthalamic nucleus; Str, striatum; Th, thalamus. (*B*) Schematic diagram depicting the hypothesized activity change over time (t) in the thalamocortical projection (Th/Cx) following the sequential inputs through the hyperdirect cortico-subthalamo-pallidal (*middle*) and direct cortico-striato-pallidal (*bottom*) pathways. (Modified from Nambu et al., 2002.)

Second, the output from the STN is excitatory, whereas the output from the striatum is inhibitory. Third, the excitatory route through the STN is faster than the inhibitory route through the striatum (Nambu et al., 2000). Finally, the STN projection to the GPi is divergent, and the striatal projection is more focused (Parent & Hazrati, 1993). Thus, the two disynaptic pathways from cerebral cortex to the basal ganglia output nuclei, the GPi and SNpr, provide fast, widespread, divergent excitation through the STN, and slower, focused, inhibition through the striatum. Because the outputs of the GPi and the SNpr are thought to be inhibitory (but see potential evidence for the contrary reviewed in Graybiel, 2005), this arrangement would result in focused facilitation and surround inhibition of basal ganglia thalamocortical targets.

In this scheme, the tonically active inhibitory output of the basal ganglia acts as a "brake" on motor control circuits of the cerebral cortex and brain stem. When a movement is initiated by a particular motor pattern generator, basal ganglia output neurons projecting to competing generators increase their firing rate, thereby increasing inhibition and applying a "brake" on these generators. Other basal ganglia output neurons projecting to the generators that are involved in the desired movement decrease their discharge, thereby removing tonic inhibition and releasing the "brake" from the desired motor patterns. Thus, the intended movement is enabled, and competing movements are prevented from interfering with the desired one.

The anatomical arrangement of STN and striatal inputs to the GPi and SNpr form the basis for a functional center-surround organization as shown in figure 39.5. When a voluntary movement is initiated by cortical mechanisms, a separate signal is sent to the STN, exciting it. The STN projects in a widespread pattern and excites the GPi. The increased GPi activity produces inhibition of thalamocortical motor mechanisms. In parallel to these pathways through the STN, signals are sent from all areas of the cerebral cortex to the striatum. The cortical inputs are transformed by integrative circuitry in the striatum to a focused, context-dependent output that inhibits specific neurons in the GPi. The inhibitory striatal input to the GPi is slower than the excitatory STN input, but it is more powerful. The resulting focally decreased activity in the GPi selectively disinhibits the desired thalamocortical motor circuits. Indirect pathways from the striatum to the GPi (striatum → external pallidum (GPe) → GPi and striatum → GPe → STN → GPi) result in further focusing of the output. The net result of basal ganglia activity during a voluntary movement is the inhibition ("braking") of competing motor patterns and focused facilitation (releasing the "brake") from the selected voluntary movement pattern generators.

This scheme provides a framework for understanding both the pathophysiology of parkinsonism and involuntary movements (Young, Albin, & Penney, 1989; Mink, 1996, 2003; Goldberg et al., 2002). Different involuntary movement disorders such as parkinsonism, chorea, dystonia, and

tic disorders result from different abnormalities in these basal ganglia circuits. Loss of the dopamine-containing nigrostriatal input to the striatum results in a loss of normal pauses of GPi discharge during voluntary movement. Hence, there is excessive inhibition of motor pattern generators and ultimately bradykinesia (Goldberg et al., 2002). Furthermore, loss of dopamine results in abnormal synchrony of GPi neuronal discharge and loss of the normal spatial and temporal focus of GPi activity (Filion, Tremblay, & Bedard, 1989; Raz, Vaadia, & Bergman, 2000; Goldberg et al., 2002). Large lesions of the GPi or the SNpr disinhibit both desired and unwanted motor patterns, leading to inappropriate activation of competing motor patterns, but normal generation of the wanted movement. Thus, lesions of GPi lead to cocontraction of multiple muscle groups and difficulty turning off unwanted motor patterns, similar to what is seen in dystonia, but they do not affect movement initiation (Mink & Thach, 1991).

Lesions of SNpr produce unwanted saccadic eye movements that interfere with the ability to maintain visual fixation but do not impair the initiation of voluntary saccades (Hikosaka & Wurtz, 1985). Lesions of putamen may result in dystonia due to the loss of focused inhibition in GPi (Mink, 2003). Lesions of the STN produce continuous involuntary movements of the contralateral limbs (hemiballism or hemichorea) (Mink, 2003). Despite the involuntary movements, voluntary movements can still be performed. Although structural lesions of putamen, GPi, SNpr, or STN produce certain types of unwanted movements or behaviors, they do not produce tics. Tics are more likely to arise from abnormal activity patterns, most likely in focal zones in the striatum (Flaherty & Graybiel, 1994; Canales & Graybiel, 2000; Mink, 2003).

The notion that the basal ganglia affect motor pattern generators helps to account for these motor disorders. But as we noted, the basal ganglia, via the thalamus, also project to a large part of the prefrontal cortex and to limbic structures. It has been proposed that these circuits act as "cognitive pattern generators" (Graybiel, 1997). Through these circuits, with similar functional structures as those considered to be motor circuits, the basal ganglia can affect cognition, planning, executive function, and our emotional lives. Some neuropsychiatric disorders, such as OCD, Tourette syndrome, and even schizophrenia, may have their origin in such cortico-basal ganglia circuits (Swerdlow & Koob, 1987; Graybiel, 1997, 2008). Possibly related to these findings is evidence that emotional distress and anguish, or irrepressible laughter and hilarity, can be evoked by deep-brain stimulation in or near the substantia nigra and the subthalamic nucleus (Bejjani et al., 1999; Krack et al., 2001; Schupbach & Agid, 2008).

The selection of which response to make is a huge job, probably engaging much of the neocortex and other brain regions, but the basal ganglia may strongly influence such selections on the basis of recognition of context, assignment of salience, and expectancy of outcome, which are discussed below. Specifically, which behaviors are selected for facilitation and which are inhibited may be based on a winner-take-all model that is established in the striatum (Dominey et al., 1995; Beiser & Houk, 1998; J. E. Brown et al., 1999; Redgrave, Prescott, & Gurney, 1999; Gillies & Arbuthnott, 2000; Frank, Loughry, & O'Reilly, 2001b; Kitano et al., 2001; Doya, 2002).

Selection is likely to take place via the compartmentally organized input-output organization of the striatum and the interaction of these compartmentalized circuits with neuromodulatory circuits (Graybiel & Ragsdale, 1978; Malach & Graybiel, 1986; Flaherty & Graybiel, 1994). If there is sufficient simultaneous activity of convergent inputs to a subgroup of striatal medium spinal neurons (Schneider & Lidsky, 1981; Flaherty & Graybiel, 1991; Kincaid, Zheng, & Wilson, 1998), they will "win," and if they are striatal neurons leading into indirect pathway, they can inhibit output neurons in the GPi and the SNpr, leading to facilitation of thalamocortical neurons. Simultaneous suppression of competing responses in this scenario would be mediated by the hyperdirect and indirect pathways. Thus, competition occurs at both the input (striatal) and output (GPi/SNpr) levels of basal ganglia circuits (Filion, Tremblay, & Bédard, 1988; Flaherty & Graybiel, 1994; Mink, 2001). This view has been adapted to account for imaging data in OCD patients (Rauch et al., 2001) and some basal ganglia models (Beiser & Houk, 1998; J. E. Brown et al., 1999; Gillies & Arbuthnott, 2000), but remains controversial.

There is substantial evidence that selection at the striatal level depends on learning mechanisms (Graybiel, 1995, 2005, 2008). Thus behavioral selections can be influenced by experience, and the striatum—and therefore the rest of the basal ganglia—participate strongly in the adaptive control of motor and cognitive behaviors. This idea originated with the recognition that corticostriatal inputs (and other inputs to the striatum) are modular, as was noted above, and that the cortical inputs show modular divergence but then can show reconvergence at the next stage of the basal ganglia circuit, within the pallidum (figure 39.7a) (Flaherty & Graybiel, 1994; Graybiel, Aosaki, Flaherty, & Kimura, 1994; Parthasarathy & Graybiel, 1997). This pattern resembles a mixture of experts learning architectures (e.g., Jacobs, Jordan, Nowlan, & Hinton, 1991), whereby information can be distributed divergently to an intermediate layer of the network, and then be gated and recombined at an output layer (figure 39.7b) (Graybiel, 1998). The dopaminergic input to the striatum could be one strong gating mechanism. Dopamine-containing inputs carry signals related to predicted reward and saliency to striatal neurons, and they are arranged anatomically to maximize their ability to gate

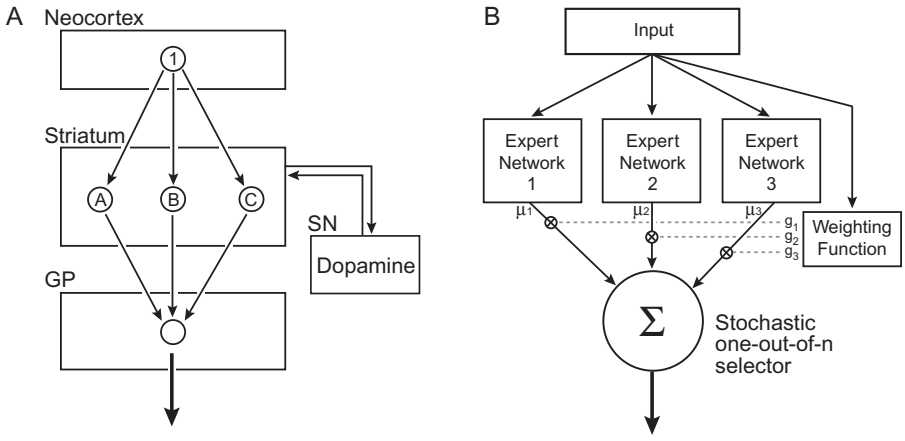


FIGURE 39.7 (*A*) Divergent-reconvergent processing of signals through cortico-basal ganglia pathways. Divergence of cortical input to modules (matrisomes A, B, C) occurs at the level of the striatum. In the globus pallidus (GP), information is reconverged, resulting in the remapping of the cortical output. The network is modulated by dopamine-containing inputs from the substantia nigra (SN). (Modified from Graybiel et al., 1994.) (*B*) Mixture-of-experts learning network model. (Modified from Jacobs et al. 1991.) Note the similarity of the models in *A* and *B*.

corticostriatal information flow (A. D. Smith & Bolam, 1990; Bolam, Hanley, Booth, & Bevan, 2000). In addition, dopamine is critically involved in long-term potentiation and depression in the striatum (Reynolds, Hyland, & Wickens, 2001; Wise, 2004; Calabresi, Piconi, Tozzi, & Di Filippo, 2007; Tang, Pawlak, Prokopenko, & West, 2007).

Many movement disorders resulting from diseases that affect the basal ganglia can be understood as disorders of response selection and inhibition. These include disorders characterized by paucity of movement, such as Parkinson's disease (Mink, 1996; Goldberg et al., 2002), and disorders characterized by excessive involuntary movements, such as chorea, dystonia, or tics (Mink, 1996, 2003; Sato et al., 2008). Notably, the treatment of Parkinson's disease by STN deep brain stimulation can improve both response selection and inhibition (Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003).

In parallel with these movement disorders, neuropsychiatric disorders can also result from impaired response selection or inhibition. Inappropriate facilitation or impaired inhibition may lead to the cognitive and motor intrusions, inflexibility, repetitiveness, and overt cognitive and motor stereotypic responses that occur in OC-spectrum disorders (Graybiel & Rauch, 2000; Leckman, 2002; Graybiel, 2008). Functional imaging studies of individuals with OC-spectrum disorders indicate increased activity in the caudate nucleus combined with increased activity in the cingulate and orbitofrontal cortices (see Rauch et al., 2001). Moreover, symptom provocation in OCD patients further increases the activity of these regions (Breiter et al., 1996), and the increased activation can be lessened by treatment of the symptoms (Baxter et al., 1992; Schwartz, Stoessel, Baxter, Martin, & Phelps, 1996; Lazaro et al., 2008). This dynamic modulation supports the idea that the basal ganglia and anterior cingulate/orbitofrontal cortical regions are important in response selection and attentional shifting.

Selecting which action to perform is critical for normal behavior. But when particular actions (or thoughts) are selected over and over again, the repetitiveness can signal the occurrence of syndromes such as OC-spectrum disorders or other disorders in which behavioral stereotypies occur. There is some evidence that the repetitiveness of action selection can be controlled independently of which action is selected. That is, different actions can be selected but, when selected, each is repetitively selected. In both rodents and primates, a specific modular pattern of neuronal activation in the striatum is highly predictive of the stereotypies induced by psychomotor stimulants: Activity in striosomes is greater than activity in the surrounding matrix regardless of which particular actions are being repeated—that is, regardless of which have been selected (figure 39.8) (see Canales & Graybiel, 2000; Saka, Goodrich, Harlan, Madras, & Graybiel, 2004). This is interesting, because anatomical work in the primate suggests that striosomes receive differentially strong input from parts of the anterior cingulate and orbitofrontal cortex. In the human, as was noted above, these are cortical regions that are abnormal in OCD patients and in addictive states (for a review, see Graybiel, 2008). Modular patterns of striatal activation have also been invoked to account for focal tics and repetitive actions in Tourette syndrome (Mink, 2001). In this case, overactivity in particular modules (matrisomes) is thought to be involved in the "selection" of the repeated behavior (figure 39.7a). Thus both striosomes and matrisomes could contribute to disorders of action selection and behavioral switching and could be important for the normal discharge of these complex functions.

Impulse-control disorders may relate to impaired response inhibition. This has become an area of substantial

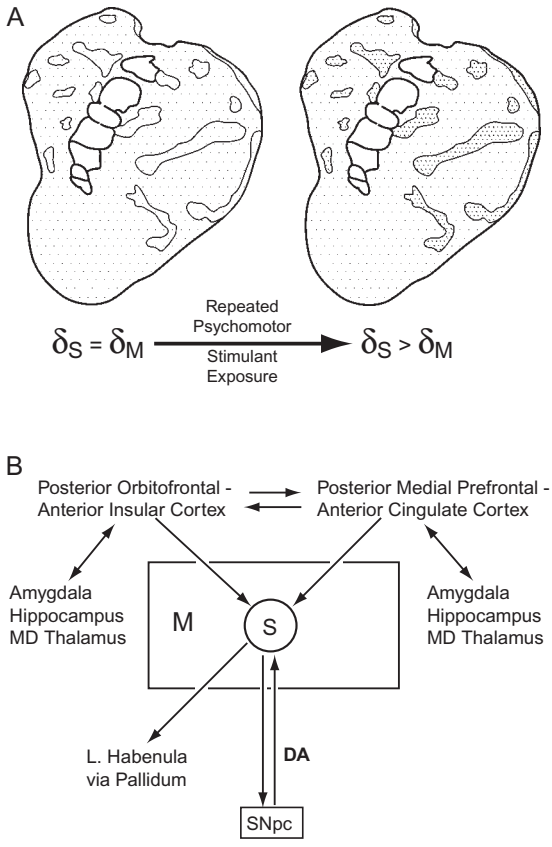


FIGURE 39.8 (*A*) Schematization of neuronal activity mapped in the caudate nucleus and putamen of the squirrel monkey in response to either single (*left*) or repeated (*right*) exposure of the monkey to psychomotor stimulants. The activity measure is the average density of striatal neurons expressing early-genes in response to the drug treatment, calculated separately for the striosome (δS) and matrix (δM) compartments. The single dose of the psychomotor stimulant induces only low levels of behavioral stereotypy and little predominance of striosomal activation. By contrast, repeated exposure to the psychomotor stimulant induces high levels of behavioral stereotypy and sharply increased striosome predominance. (*B*) Schematization of the major connections of the striosomes. The central rectangle represents the striatum with its matrix (M) and striosomal (S) compartments. Abbreviations: MD, mediodorsal nucleus; SNpc, substantia nigra, pars compacta; DA, dopamine. (Modified from Graybiel, 1997.)

interest recently in relation to Parkinson's disease. Patients with Parkinson's disease are commonly treated with dopamine replacement therapy (levodopa) or with direct dopamine receptor agonists. A variety of impulse control disorders, including compulsive gambling and excessive risk taking ("punding"), have been described in patients with Parkinson's disease who are treated with these dopaminergic medications (Dodd et al., 2005; Pontone, Williams, Bassett, & Marsh, 2006; Weintraub et al., 2006). As was discussed above, the STN is thought to play a central role in response inhibition (Aron & Poldrack, 2006; Eagle et al., 2008), and it is thought that impulse control disorders in Parkinson's disease may relate to impairment of STN-mediated response inhibition. Indeed, STN deep brain stimulation may cause impaired impulse control despite improvement of other aspects of movement (Frank, 2006; Winstanley, Eagle, & Robbins, 2006).

ATTENTION AND ASSIGNMENT OF SALIENCE The attention-salience assignment model of the basal ganglia suggests that the outputs of the basal ganglia are influential in modulating movement because they can influence attention to stimuli and because they have the capacity to assign salience to stimuli. This idea is strongly supported by work on the dopamine-containing inputs to the basal ganglia, which carry signals related to reinforcement probability, salience, and expectation of reinforcement (Schultz, Dayan, & Montague, 1997; Berridge & Robinson, 1998; Doya, 2002; Glimcher, 2003; Daw, Niv, & Dayan, 2005; Niv, Duff, & Dayan, 2005; Niv, Joel, & Dayan, 2006; Schultz, 2007; Graybiel, 2008). We note, however, that several other systems that could have this function also project to the basal ganglia. These include the locus coeruleus/norepinephrine system (projecting especially strongly to the ventral striatum), the serotonergic raphe system, the intralaminar thalamic nuclei, and other structures such as the amygdala.

Clinical studies also have repeatedly implicated the basal ganglia in attentional control (Mesulam, 2000), and, in modern formulations of this idea, the basal ganglia are particularly singled out as being important for "attention to action" (Jueptner, Stephan, et al., 1997). Imaging studies indicate that cortico-basal ganglia circuit dysfunction in Parkinson's disease may account for the marked attentional problems suffered by Parkinson's disease patients (see Saint-Cyr, 2003). There is, in addition, evidence for dysfunction of corticocortical connections linking the supplementary motor area and premotor areas (Rowe et al., 2002). This dysfunction at the cortical level could itself be related to abnormal basal ganglia influences on these cortical areas (Brooks, 1997; Samuel et al., 1997). It should be clear, however, that "attention" is a broad concept and, in the context of cortico-basal ganglia loops, includes functions ranging from saliency signals modulating signal-to-noise ratios to motor readiness (Denny-Brown & Yanagisawa, 1976; Robbins & Everitt, 1992; Aosaki, Graybiel, & Kimura, 1994; L. L. Brown, Schneider, & Lidsky, 1997; Jog et al., 1999; Barnes et al., 2005). Considered in this way, attentional deficits in Parkinson's disease could lead to bradykinesia (slowness of movement), bradyphrenia (slowness of thought), and abulia (a cardinal sign of anterior striatal dysfunction in which a profound inertia of psychomotor response initiation occurs).

Patients with Parkinson's disease and Huntington's disease exhibit deficits in shifting of attention, termed set-shifting (Owen et al., 1993; Georgiou, Bradshaw, Phillips, & Chiu,

1996; Bédard et al., 1998). Again, these are disorders with widespread neurodegeneration not confined to the basal ganglia proper, but in normal individuals, there are significant and selective increases in blood flow in the striatum for tasks that measure attention shifts in response to visual cues (Koski, Paus, Hofle, & Petrides, 1999).

A dramatic deficiency in attentional control is present in ADHD, in which individuals exhibit hyperactive behavior, a lack of focusing ability, and impulsivity. Functional imaging studies addressing the possible involvement of the basal ganglia in this disorder suggest that the capacity to inhibit motor activity and the capacity to sustain attention may be linked in ADHD individuals, and that these clinical measures of abnormal function are correlated with altered activity in the putamen (Teicher et al., 2000).

There now is direct evidence that activity in the striatum is important as an attentional filter in humans (McNab & Klingberg, 2008), and there is direct electrophysiological evidence in animals for striatal representations that emphasize salient events and deemphasize others (Jog et al., 1999; Barnes et al., 2005). Electrophysiological studies in primates also support the idea that the basal ganglia are part of forebrain attentional systems. Explicit tests of attentional shifting suggest that many striatal projection neurons fire for shifts in attention that are unaccompanied by overt movements (Kermadi & Boussaoud, 1995; Boussaoud & Kermadi, 1997). An instructive example comes from studies of striatal interneurons called tonically active neurons (TANs), which are broadly distributed through the caudate nucleus and putamen (figure 39.9). These neurons modify their responses to sensory stimuli depending on the saliency of the sensory stimuli. The salience can be unconditional (e.g., a loud, unexpected sound makes them respond) or can be built up through conditioning by pairing the sensory cues with positive or negative reinforcements (Aosaki, Tsubokawa, et al., 1994; Apicella, 2002; Blazquez et al.,

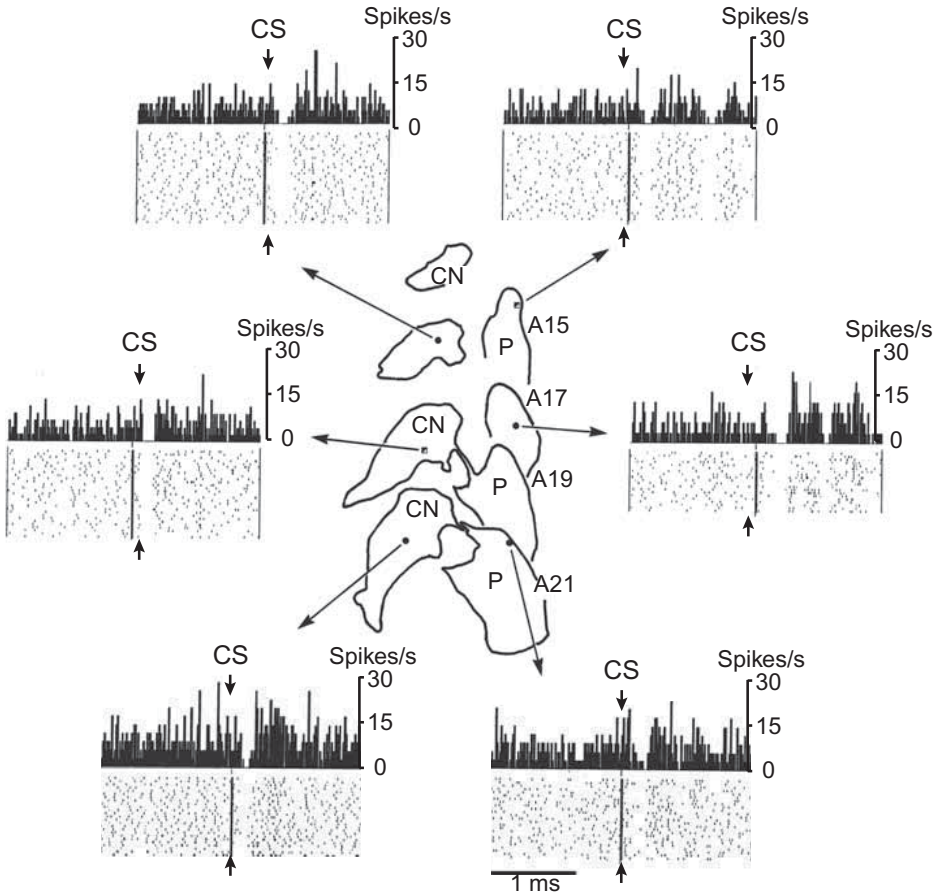


FIGURE 39.9 The responses of tonically active neurons (TANs) of the macaque monkey striatum in response to conditioned stimuli (CS: clicks or light-emitting diodes) in a simple behavioral conditioning paradigm in which the monkey receives liquid rewards following delivery of the CS. The neurons acquire responses to the cues associated with the rewards (see pauses in activity). The responses of six representative TANs recorded at the illustrated sites (*black dots or squares*) are shown in raster plots and spike histograms, and the anteroposterior (AP) sites at which they were recorded are shown in diagrams. Note the widespread, coherent appearance of the response, suggesting that these interneurons might serve as a temporal binding mechanism across cortico-basal ganglia loops. (Adapted from Graybiel et al., 1994.)

2002). The reward/saliency signals are partly dependent on inputs from the dopamine-containing neurons of the substantia nigra (Aosaki et al., 1994). But they depend also on inputs from the intralaminar nuclei of the thalamus (Matsumoto, Minamimoto, Graybiel, & Kimura, 2001) and probably the neocortex as well. Because the TANs are widely distributed local network neurons, and tend to have synchronous responses (Graybiel, Aosaki, Flaherty, & Kimura, 1994; Raz, Feingold, Zelanskaya, Vaadia, & Bergman, 1996; Blazquez, Fuji, Kojima, & Graybiel, 2002), they could coordinate together activity in functionally distinct cortico-basal ganglia loops to achieve sensorimotor and cognitive binding (figure 39.10) (see Graybiel et al., 1994; see Graybiel, 1997).

This example gives an idea of how signals from many different brain regions could ultimately lead to salience signaling in basal ganglia networks and contribute to motor and cognitive attention. Remarkably, it has been estimated that the population activity of even a small number of these interneurons can accurately predict ongoing behavioral events (Blazquez et al., 2002). This means that the intrinsic circuitry of the striatum has within it a signal that is proportional to behavioral outcome—exactly what is needed to develop a forward model for behavioral control (Blazquez et al., 2002).

Learning and Adaptive Control of Behavior The idea that the basal ganglia are sites for learning has strong support from experimental work in animals and increasing support from imaging and other work in humans. Commonly, attempts to formulate how the basal ganglia could contribute to learning involve comparing the basal ganglia to the hippocampus, or comparing them to the cerebellum (Packard & Knowlton, 2002; Doyon, Penhune, & Ungerleider, 2003). Helpful as such comparisons may be, they are not enough to define the type of neural processing that occurs in basal ganglia networks and cortico-basal ganglia circuits during behavioral learning. As noted below, new techniques are now beginning to let investigators approach this issue directly.

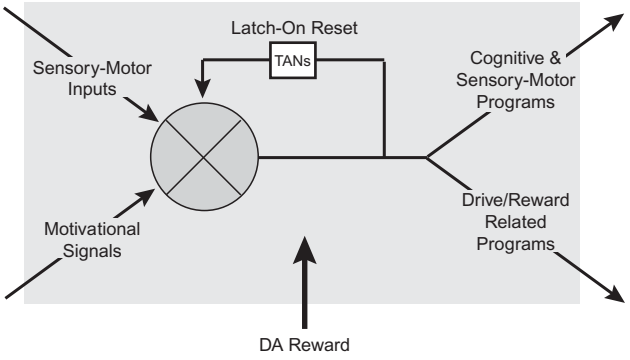


Figure 39.10 Diagram illustrating the hypothesis that the striatum acts as a dynamic modulator of cognitive and motor programs and that striatal interneurons, including TANs, function as part of the plastic neural mechanism underlying this dynamic modulation. (Adapted from Graybiel et al., 1994.)

Studies in rat, monkey, and human suggest that the basal ganglia mediate a particular type of learning and memory: stimulus-response (S-R) learning, in which learning proceeds by trial and error and performance improves according to the sensory feedback obtained as a result of the response (Packard & McGaugh, 1992; McDonald & White, 1994). Evidence for this has led to the notion that the basal ganglia are important for habit or skill learning (Graybiel, 1995, 2005, 2008; Packard & Knowlton, 2002). The function of the basal ganglia in feedback (S-R) learning appears to be highly conserved. In birds, the anterior forebrain pathway (AFP) is thought to be analogous to certain cortico-basal ganglia circuits in mammals, and it has been shown that this AFP pathway is critical to bird song learning (Brainard & Doupe, 2000; Amin, Doupe, & Theunissen, 2007; Calaminus & Hauber, 2007; Aronov, Andalman, & Fee, 2008).

Experimental psychologists working with rats have amassed strong evidence that the caudoputamen (dorsal striatum) is necessary for both the acquisition and the expression of S-R associations and memory and for "win-stay" learning in which the animal repeats the behavior that led to reward (Packard, Hirsh, & White, 1989; Packard & McGaugh, 1992, 1996; McDonald & White, 1994). This behavior is contrasted with "win-shift" behavior involving explicit memory of the context of the behavior. However, the situation may be more than one of different basal ganglia loops participating in different aspects of learning. For example, performance on S-R learning tasks suffers in rats with lateral (sensorimotor) striatal lesions; but in rats with medial striatal lesions, performance suffers on tasks similar to those requiring hippocampal function, for example, spatial navigation (Devan, McDonald, & White, 1999; Devan & White, 1999; Packard & Knowlton, 2002; Yin & Knowlton, 2006). Moreover, there is good reason to think that in most of the tasks that are used in such rodent experiments, explicit awareness of the associations (e.g., place learning) could occur, engaging the hippocampus. Some studies suggest that in such contexts, the hippocampus may operate during an early, explicit stage of learning and that the striatum may then take over (or at least be more critical) when the task is repeated to the point at which the animal can perform the task without explicit knowledge (McDonald & White, 1994; Packard & McGaugh, 1996). Thus even with damage to the striatum, habit learning could be partly intact because learning strategies based on hippocampal function can partly compensate for the deficient recruitment of the basal ganglia (Packard & Knowlton, 2002). Interestingly, evidence in these rodent studies suggests that the striatum and hippocampus can compete with each other, so that a lesion of one system

may actually facilitate learning mediated by the other system. For example, a deficit in spatial learning strategy following hippocampal damage can improve performance of S-R learning (Packard et al., 1989; McDonald & White, 1993; Schroeder, Wingard, & Packard, 2002).

The electrophysiological recordings that are made as animals learn association tasks demonstrate remarkable plasticity in the responses of striatal neurons as the animals learn (Jog et al., 1999; Barnes et al., 2005; for reviews, see Graybiel, 2005, 2008; Pasupathy & Miller, 2005). If monkeys learn a task and then the task requirements are reversed, striatal neurons are quick to acquire the new (reversed) association (Pasupathy & Miller, 2005). In association maze-learning tasks, there are dramatic changes in the patterns of task-related activity of striatal neurons in the sensorimotor striatum (Jog et al., 1999; Barnes et al., 2005). These changes at the neuronal level have been likened to the explore-exploit behaviors that are delineated in reinforcement learning models (Sutton & Barto, 1998; Barnes et al., 2005). Remarkably, just such models have been invoked to account for the acquisition of song in avian species that learn their songs (Doya & Sejnowski, 1995). Interestingly, as the task-related neurons change their firing patterns, other neurons that do not exhibit phasic spike activity in relation to the task gradually become nearly silent.

In imaging studies of human performance, activation of the basal ganglia has been repeatedly found to accompany motor skill learning. Skill learning can be broken down into a number of phases but, as in learning a sport, requires practice, S-R (feedback) learning, and consolidation (Brashers-Krug, Shadmehr, & Bizzi, 1996; Karni et al., 1998; Hikosaka et al., 1999; Ungerleider, Doyon, & Karni, 2002; Saint-Cyr, 2003). Once learned, the sequence of movements can be carried out seemingly effortlessly with the same or similar effector groups used during practice. Many studies have employed the serial reaction time (SRT) task to study simple human motor skill learning (Grafton et al., 1995; Willingham, Salidis, & Gabrieli, 2002). For example, subjects can be asked to press a series of buttons in an order instructed by target lights that appear either in a random sequence or in a predetermined, repeated sequence. With practice, the subjects become faster, especially with the repeated sequences. If the subject is told about the sequence beforehand, the reaction time advantage for the repeating sequence is thought to occur by virtue of declarative learning, but to involve nondeclarative, implicit learning if the subject does not know about the repeating sequence (or is distracted by a second task). Positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies have demonstrated heightened activation in the putamen, along with a network of cortical areas, in the implicit condition (Grafton et al., 1995; Hazeltine, Grafton, & Ivry, 1997; Willingham et al., 2002), and in some explicit conditions as well (Willingham et al., 2002; Doyon et al., 2003; Doyon & Benali, 2005).

Learning and performing a sequence of finger movements by trial and error with auditory feedback evokes activation of the striatum also, both in learning of a new sequence and in the execution of a pre-learned sequence (Jenkins, Brooks, Nixon, Frackowiak, & Passingham, 1994). The acquisition phase favors more anterior activation (caudate nucleus, anterior putamen) by comparison with performance of a pre-learned sequence (Jueptner, Frith, Brooks, Frackowiak, & Passingham, 1997, Jueptner, Stephan, et al., 1997). Similar anterior-to-posterior shifts also occur in the frontal cortex during learning (Jueptner, Frith, et al., 1997). Attention to action may in part underlie the shift. When subjects attend to their next action in a pre-learned (automized) sequence, the caudate nucleus, but not the (more posterior) putamen, exhibits differentiated activation (Jueptner, Frith, et al., 1997). Quite similar anterior-posterior gradients have been found in primates (Miyachi, Hikosaka, Miyashita, Karadi, & Rand, 1997; Nakamura, Sakai, & Hikosaka, 1998, 1999; Hikosaka et al., 1999).

If we think back to the anatomy of cortico-basal ganglia loops, we can see that these and other studies (Shadmehr & Brashers-Krug, 1997; Honda et al., 1998; Jueptner & Weiller, 1998; Karni et al., 1998; Peterson et al., 1998) suggest that the acquisition of motor skills probably engages the activity of a number of corticostriatal loop systems, and that which loops are engaged changes during different stages of learning, from the first learning of the basic structure of the task (its "rules" or constraints) to an eventual engagement of particular muscle groups in sequence without conscious calling up of the single parts of the behavior. The early stages activate cortico-basal ganglia loops in which the caudate nucleus and anterior putamen participate, and later stages activate putamen-based loops. Interestingly, contrary to the activation of the putamen in motor skill learning, perceptual skill learning (e.g., a mirror reading task) is linked to activation of the caudate nucleus (Poldrack & Gabrieli, 2001). The cerebellum is also activated in such tasks. One interesting idea is that early phases of S-R sequential learning engage spatial coordinate frames and later phases motor coordinate frames (Hikosaka et al., 1999).

More cognitive versions of S-R learning tasks, requiring implicit learning by feedback of probabilistic classifications, also differentially activate the caudate nucleus (Saint-Cyr, Taylor, & Lang, 1988; Knowlton, Squire, & Gluck, 1994; Poldrack et al., 1999). The medial temporal lobe, by contrast, is activated when such tasks are acquired through observation (paired-association tasks) rather than through guessing and learning by trial and error (Poldrack & Gabrieli, 2001). Supporting the idea of antagonistic

activity of striatal and hippocampal systems raised by studies in experimental animals, imaging studies in human subjects demonstrate deactivation of the medial temporal lobe during acquisition of the feedback-based task. The activities of the caudate nucleus and of the medial temporal are negatively correlated (Poldrack & Gabrieli, 2001). Patients with Parkinson's disease and Huntington's disease perform more poorly in the feedback-based probabilistic classification task than do patients with localized frontal lobe lesions, suggesting that it is not only a dysfunction of the frontal part of the frontobasal ganglia loop, but, more likely, deficits in neuronal processing in the striatum itself that lead to learning deficits in Parkinson's disease patients (Knowlton, Mangels, & Squire, 1996).

Impairments in motor and perceptual skill learning have been demonstrated in patients with Huntington's disease and Parkinson's disease (Martone, Butters, Payne, Becker, & Sax, 1984; Heindel, Butters, & Salmon, 1988; M. A. Smith, Brandt, & Shadmehr, 2000). Patients with OCD have deficits in performing the implicit form of SRT tasks when a second task is introduced (Deckersbach et al., 2002), and they fail to exhibit activation of the striatum during the acquisition of SRT tasks (Rauch et al., 2001).

We have concentrated on the dorsal striatum (the caudate nucleus and putamen), but evidence suggests that the ventral striatum is also critical to reinforcement-based learning, together with its dopaminergic input from the ventral tegmental area. For example, neurons of the ventral striatum can apparently keep track of how close a monkey is to receiving reward (Bowman, Aigner, & Richmond, 1996; Shidara, Aigner, & Richmond, 1998; Rolls, 1999). Cues related to reward, and reward itself, can actuate these neurons (Ito, Dalley, Robbins, & Everitt, 2002; Phillips, Stuber, Heien, Wightman, & Carelli, 2003; Tanaka et al., 2004; Zald et al., 2004; Taha, Nicola, & Fields, 2007; Lansink et al., 2008). Lesions of the ventral striatum can block acquisition of approach maze task problems (Atallah, Lopez-Paniague, Rudy, & O'Reilly, 2007).

The striosomes of the dorsal striatum, by virtue of their connections with many of the same brain structures as the ventral striatum, are likely also to be important in the learning and execution of rewarded tasks (Aosaki, Kimura, & Graybiel, 1995; White & Hiroi, 1998). As the dorsal striatum and ventral striatum are believed to participate in different forms of learning (nondeclarative and declarative, respectively), it is possible that the reward evaluation function of the ventral striatum is taken over by striosomes in nondeclarative learning. As we noted above, activity in striosomes is correlated with maladaptive, perseverative responses following psychomotor stimulant exposure, raising the possibility that they could be involved also in stereotypic behaviors in OC-spectrum disorders.

Chunking of action repertoires as a common theme for basal ganglia function

We have considered here three categories of hypotheses about the functions of the basal ganglia, ranging from the selective facilitation and inhibition of movements (or thoughts), to the assignment of saliences and attention to stimuli, to behavioral learning, especially feedback-based learning. There are other behavioral categories that should also be considered, including sequencing of movements or cognitive acts, scaling or timing of these acts, and preparing for the next movement or thought. But regardless of the behavioral categorization, we still must remember that the basal ganglia are embedded in circuits that engage the thalamus and cerebral cortex and other sites as well. How can we learn what part of any function to attribute to the basal ganglia, and what part to other structures in these basal ganglia-based circuits?

One important recent finding from primate physiology is that identified corticostriatal neurons in the motor cortex have quite different response properties than even very nearby motor cortex neurons projecting to the spinal cord (Turner & DeLong, 2000). In trained monkeys, at least, the responses of the neurons seem tuned to very discrete contexts, and they are nearly all direction-selective. This finding suggests that the information reaching the striatum is not an exact copy (efference copy or corollary discharge) of the motor command sent to the spinal cord. But it could be, for example, that cortical inputs to the subthalamic nucleus (hyperdirect pathway) are; this is not yet known. There is also suggestive evidence that inputs from the motor cortex tend to activate striatal neurons of the indirect pathway more than those of the direct pathway (Berretta, Parthasarathy, & Graybiel, 1997; Parthasarathy & Graybiel, 1997; Lei, Jiao, Del Mar, & Reiner, 2004), and that the reverse is true for thalamostriatal inputs (Y. Smith, Bevan, Shink, & Bolam, 1998). Even these two examples indicate that our understanding of cortico-basal ganglia networks is still primitive.

Another approach to the circuit issue has been to record in the striatum as animals undergo training in behaviors that are thought to require striatal function—as was discussed above, procedural, S-R, habit or "win-stay" learning. For example, in the experiment illustrated in figure 39.11, rats were trained to run down a simple T-maze to obtain reward at one or the other of the end arms, and conditional auditory cues were given during the maze run to tell the animal which arm was baited (Jog et al., 1999). Each day during training, physiological recordings of the firing of ensembles of striatal neurons were made with tetrodes chronically implanted in the sensorimotor sector of the striatum. As shown in figure 39.11, there was a dramatic change in response patterns of striatal neurons during learning. Responses during the turn

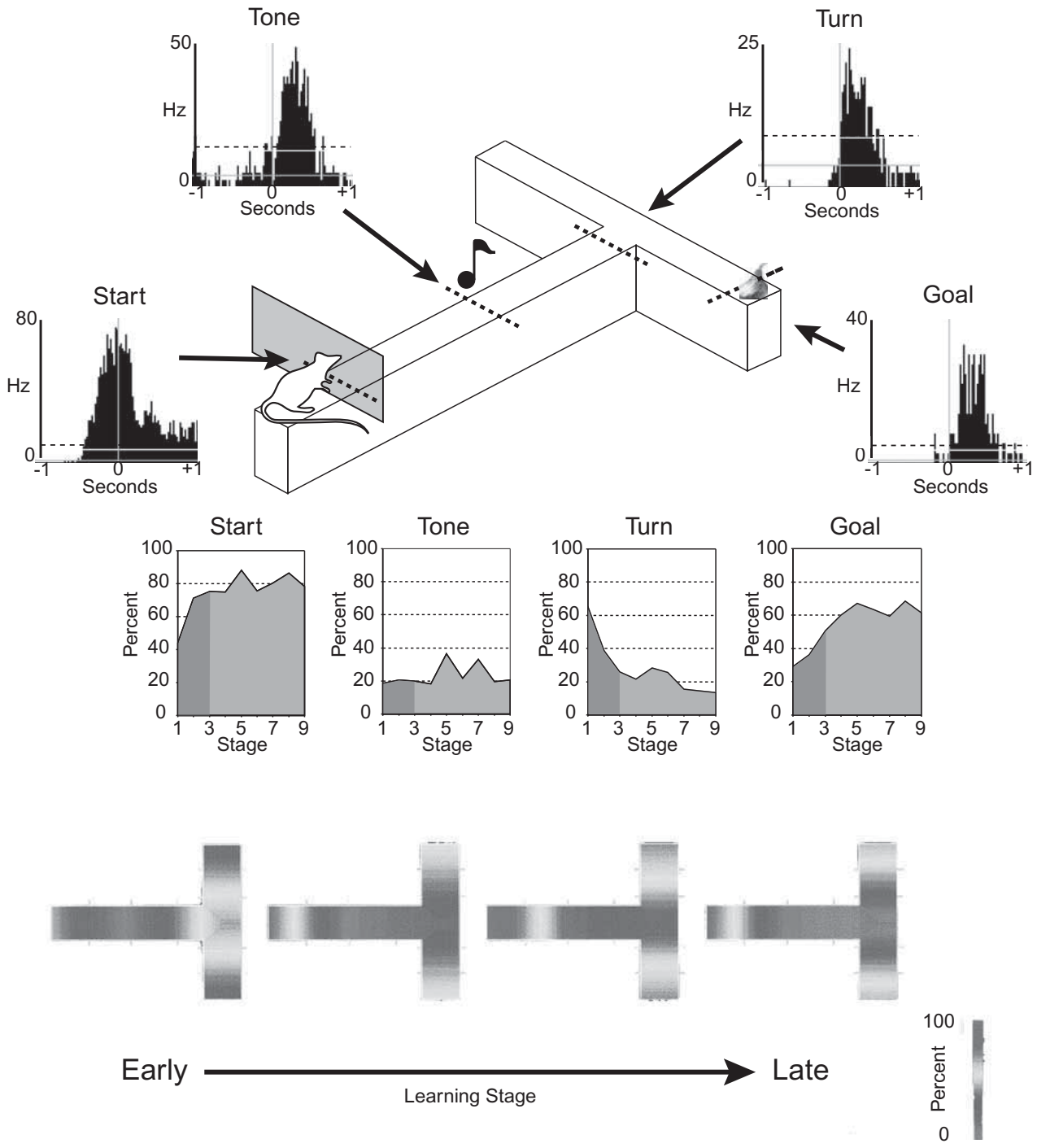


FIGURE 39.11 Event-related ensemble activity of neurons in the dorsolateral striatum of rats recorded during the acquisition and performance of an auditory conditional turning task in a T-maze. Perievent histograms displayed around the T-maze show examples of the activities of single striatal neurons in relation to start, tone, turn, and goal events. Plots below the maze illustrate the reorganization of task-related activity patterns of the striatal neurons that occurs during the acquisition of the task. The behavioral criterion for acquisition was at stage 3. Color plots at bottom illustrate schematically the gradual changes in the response profiles of the striatal neurons during the course of behavioral learning. (Modified from Jog et al., 1999; Graybiel and Kubota, 2003.) (See color plate 52.)

part of the task declined, but responses at the start and end of the task increased. In this experiment, at least, what seemed to happen is that the task boundaries became emphasized by ensemble activity in the striatum, and in-between parts of the task became less prominently represented (Jog et al., 1999; Graybiel & Kubota, 2003). This start and end accentuation has been replicated (Barnes et al., 2005) and has now been found in corticostriatal loops in awake behaving primates, suggesting that the representation of action boundaries could be an important aspect of encoding for cortico-basal ganglia circuits (Fujii & Graybiel, 2003).

How could this work? One idea is that, through activity in cortico-basal ganglia loops, the representations of actions that are repeated over and over again get recoded: Representations related to entire sequences of actions making up a behavior are built so that each individual element of the behavior is no longer coded in detail in the loop in question. In this case, the basal ganglia could be viewed as structures that help to remap action representations into expressible units. By analogy to memory units, these have been called chunks (Graybiel, 1998). This could potentially be a common theme tying together the apparently different categories of behavior attributed to the basal ganglia and discussed here. For example, the release-inhibition function, seen in these terms, would suggest that for automated behaviors, the behaviors are releasable (or repressible) as a whole, not necessarily element by element. For the attention and assignment of salience, cortico-basal ganglia circuits, through repetition, could be remapped so that particular salient sensory cues or "contexts" could trigger entire responses. This is, in fact, one definition of habit (James, 1890). Similarly, response selection, through repetition and learning, could be automated by such a chunking process. Finally, in this view, learning functions would be seen as critical and core functions of cortico-basal ganglia networks and, in particular, of corticostriatal networks.

REFERENCES

Albin, R. L., & Mink, J. W. (2006). Recent advances in Tourette syndrome research. *Trends Neurosci.*, *29*, 175–182.

Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends Neurosci.*, *12*, 366–375.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.*, *9*, 357–381.

Amin, N., Doupe, A., & Theunissen, F. E. (2007). Development of selectivity for natural sounds in the songbird auditory forebrain. *J. Neurophysiol.*, *97*, 3517–3531.

Aosaki, T., Graybiel, A. M., & Kimura, M. (1994). Effects of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. *Science*, *265*, 412–415.

Aosaki, T., Kimura, M., & Graybiel, A. M. (1995). Temporal and spatial characteristics of tonically active neurons of the primate's striatum. *J. Neurophysiol.*, *73*, 1234–1252.

Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A. M., & Kimura, M. (1994). Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *J. Neurosci.*, *14*, 3969–3984.

Apicella, P. (2002). Tonically active neurons in the primate striatum and their role in the processing of information about motivationally relevant events. *Eur. J. Neurosci.*, *16*, 2017–2026.

Apicella, P. (2007). Leading tonically active neurons of the striatum from reward detection to context recognition. *Trends Neurosci.*, *30*, 299–306.

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: Role of the subthalamic nucleus. *J. Neurosci.*, *26*, 2424–2433.

Aronov, D., Andalman, A. S., & Fee, M. S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science*, *320*, 630–634.

Atallah, H. E., Lopez-Paniagua, D., Rudy, J. W., & O'Reilly, R. C. (2007). Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nat. Neurosci.*, *10*, 126–131.

Barnes, T., Kubota, Y., Hu, D., Jin, D. Z., & Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, *437*, 1158–1161.

Baxter, L. R., Jr., Schwartz, J. M., Bergman, K. S., Szuba, M. P., Guze, B. H., Mazziotta, J. C., Alazraki, A., Selin, C. E., Ferng, H. -K., & Phelps, M. E. (1992). Caudate glucose metabolic rate changes with both drug and behavioral therapy of obsessive-compulsive disorder. *Arch. Gen. Psychiatry*, *49*, 681–689.

Bédard, M. -A., Agid, Y., Chouinard, S., Fahn, S., Korczyn, A. D., & Lespérance, P. (Eds.). (2003). *Mental and behavioral dysfunction in movement disorders*. Towawa, NJ: Humana.

Bédard, M. A., el Massioui, F., Malapani, C., Dubois, B., Pillon, B., Renault, B., & Agid, Y. (1998). Attentional deficits in Parkinson's disease: Partial reversibility with naphtoxazine (SDZ NVI-085), a selective noradrenergic alpha 1 agonist. *Clin. Neuropharmacol.*, *21*, 108–117.

Beiser, D. G., & Houk, J. C. (1998). Model of cortical-basal ganglionic processing: Encoding the serial order of sensory events. *J. Neurophysiol.*, *79*, 3168–3188.

Bejjani, B. P., Damier, P., Arnulf, I., Thivard, L., Bonnet, A. M., Dormont, D., Cornu, P., Pidoux, B., Samson, Y., & Agid, Y. (1999). Transient acute depression induced by high-frequency deep-brain stimulation. *N. Engl. J. Med.*, *340*, 1476–1480.

Benabid, A. L., Vercucil, L., Benazzouz, A., Koudsie, A., Chabardes, S., Minotti, L., Kahane, P., Gentil, M., Lenartz, D., Andressen, C., Krack, P., & Pollak, P. (2003). Deep brain stimulation: What does it offer? *Adv. Neurol.*, *91*, 293–302.

Bergman, H., Feingold, A., Nini, A., Raz, A., Slovin, H., Abeles, M., & Vaadia, E. (1998). Physiological aspects of information processing in the basal ganglia of normal and parkinsonian primates. *Trends Neurosci.*, *21*, 32–38.

Bergman, H., Wichmann, T., & DeLong, M. R. (1990). Reversal of experimental parkinsonism by lesions of the subthalamic nucleus. *Science*, *249*, 1436–1438.

Berretta, S., Parthasarathy, H. B., & Graybiel, A. M. (1997). Local release of GABAergic inhibition in the motor cortex induces immediate-early gene expression in indirect pathway neurons of the striatum. *J. Neurosci.*, *17*, 4752–4763.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.*, *28*, 309–369.

Bhatia, K. P., & Marsden, C. D. (1994). The behavioural and motor consequences of focal lesions of the basal ganglia in man. *Brain*, *117*, 859–876.

Blazquez, P., Fujii, N., Kojima, J., & Graybiel, A. M. (2002). A network representation of response probability in the striatum. *Neuron*, *33*, 973–982.

Bolam, J. P., Hanley, J. J., Booth, P. A., & Bevan, M. D. (2000). Synaptic organisation of the basal ganglia. *J. Anat.*, *196*(Pt. 4), 527–542.

Boussaoud, D., & Kermadi, I. (1997). The primate striatum: Neuronal activity in relation to spatial attention versus motor preparation. *Eur. J. Neurosci.*, *9*, 2152–2168.

Bowman, E. M., Aigner, T. G., & Richmond, B. J. (1996). Neural signals in the monkey ventral striatum related to motivation for juice and cocaine rewards. *J. Neurophysiol.*, *75*, 1061–1073.

Braak, H., Del Tredici, K., Rub, U., de Vos, R. A., Jansen Steur, E. N., & Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging*, *24*, 197–211.

Brainard, M. S., & Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. *Nat. Rev. Neurosci.*, *1*, 31–40.

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, *382*, 252–255.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, *5*, 49–62.

Breiter, H. C., Rauch, S. L., Kwong, K. K., Baker, J. R., Weisskoff, R. M., Kennedy, D. N., Kendrick, A. D., Davis, T. L., Jiang, A., Cohen, M. S., Stern, C. E., Belliveau, J. W., Baer, L., O'Sullivan, R. L., Savage, C. R., Jenike, M. A., & Rosen, B. R. (1996). Functional magnetic resonance imaging of symptom provocation in obsessive-compulsive disorder. *Arch. Gen. Psychiatry*, *53*, 595–606.

Brooks, D. J. (1997). Advances in imaging Parkinson's disease. *Curr. Opin. Neurol.*, *10*, 327–331.

Brown, J. E., Bullock, D., & Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibatory learning pathways to selectively respond to unexpected rewarding cues. *J. Neurosci.*, *19*, 10502–10511.

Brown, L. L., Schneider, J. S., & Lidsky, T. I. (1997). Sensory and cognitive functions of the basal ganglia. *Curr. Opin. Neurobiol.*, *7*, 157–163.

Brown, R. G., & Marsden, C. D. (1990). Cognitive function in Parkinson's disease: From description to theory. *Trends Neurosci.*, *13*, 21–29.

Calabresi, P., Picconi, B., Tozzi, A., & Di Filippo, M. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends Neurosci.*, *30*, 211–219.

Calaminus, C., & Hauber, W. (2007). Intact discrimination reversal learning but slowed responding to reward-predictive cues after dopamine D1 and D2 receptor blockade in the nucleus accumbens of rats. *Psychopharmacology (Berl.)*, *191*, 551–566.

Calzavara, R., Mailly, P., & Haber, S. N. (2007). Relationship between the corticostriatal terminals from areas 9 and 46, and those from area 8A, dorsal and rostral premotor cortex and area 24c: An anatomical substrate for cognition to action. *Eur. J. Neurosci.*, *26*, 2005–2024.

Canales, J. J., & Graybiel, A. M. (2000). A measure of striatal function predicts motor stereotypy. *Nat. Neurosci.*, *3*, 377–383.

Caplan, L. R., Schmahmann, J. D., Kase, C. S., Feldmann, E., Baquis, G., Greenberg, J. P., Gorelick, P. B., Helgason, C., & Hier, D. B. (1990). Caudate infarcts. *Arch. Neurol.*, *47*, 133–143.

Chang, C., Crottaz-Herbette, S., & Menon, V. (2007). Temporal dynamics of basal ganglia response and connectivity during verbal working memory. *NeuroImage*, *34*, 1253–1269.

Cools, R., Gibbs, S. E., MiyakaWa, A., Jagust, W., & D'Esposito, M. (2008). Working memory capacity predicts dopamine synthesis capacity in the human striatum. *J. Neurosci.*, *28*, 1208–1212.

Cools, R., Ivry, R. B., & D'Esposito, M. (2006). The human striatum is necessary for responding to changes in stimulus relevance. *J. Cogn. Neurosci.*, *18*, 1973–1983.

Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*, 1510–1512.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, *8*, 1704–1711.

Deckersbach, T., Savage, C. R., Curran, T., Bohne, A., Wilhelm, S., Baer, L., Jenike, M. A., & Rauch, S. L. (2002). A study of parallel implicit and explicit information processing in patients with obsessive-compulsive disorder. *Am. J. Psychiatry*, *159*, 1780–1782.

DeLong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends Neurosci.*, *13*, 281–289.

Denny-Brown, D., & Yanagisawa, N. (1976). The role of the basal ganglia in the initiation of movement. *Res. Publ. Assoc. Res. Nerv. Ment. Dis.*, *55*, 115–149.

Deschenes, M., Bourassa, J., Doan, V. D., & Parent, A. (1996). A single-cell study of the axonal projections arising from the posterior intralaminar thalamic nuclei in the rat. *Eur. J. Neurosci.*, *8*, 329–343.

Desmond, J. E., Gabrieli, J. D. E., & Glover, G. H. (1998). Dissociation of frontal and cerebellar activity in a cognitive task: Evidence for a distinction between selection and search. *NeuroImage*, *7*, 368–376.

Devan, B. D., McDonald, R. J., & White, N. M. (1999). Effects of medial and lateral caudate-putamen lesions on place- and cue-guided behaviors in the water maze: Relation to thigmotaxis. *Behav. Brain Res.*, *100*, 5–14.

Devan, B. D., & White, N. M. (1999). Parallel information processing in the dorsal striatum: Relation to hippocampal function. *J. Neurosci.*, *19*, 2789–2798.

Divac, I., Rosvold, H. E., & Szwarcbart, M. K. (1967). Behavioural effects of selective ablation of the caudate nucleus. *J. Comp. Physiol. Psychol.*, *63*, 184–190.

Dodd, M. L., Klos, K. J., Bower, J. H., Geda, Y. E., Josephs, K. A., & Ahlskog, J. E. (2005). Pathological gambling caused by drugs used to treat Parkinson disease. *Arch. Neurol.*, *62*, 1377–1381.

Dominey, P., Arbib, M., & Joseph, J. P. (1995). A model of corticostriatal plasticity for learning oculomotor associations and sequences. *J. Cogn. Neurosci.*, *7*, 311–336.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Net.*, *15*, 495–506.

Doya, K., & Sejnowski, T. J. (1995). A novel reinforcement model of birdsong vocalization learning. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), *Advances in neural information processing systems* (Vol. 7, pp. 101–108). Cambridge, MA: MIT Press.

Doyon, J., & Benali, H. (2005). Reorganization and plasticity in the adult brain during learning of motor skills. *Curr. Opin. Neurobiol.*, *15*, 161–167.

Doyon, J., Penhune, V., & Ungerleider, L. G. (2003). Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*, *41*, 252–262.

Dubois, B., Pillon, B., & Agid, Y. (1992). Deterioration of dopaminergic pathways and alterations in cognition and motor functions. *J. Neurol.*, *239*(Suppl 1), S9–S12.

Eagle, D. M., Baunez, C., Hutcheson, D. M., Lehmann, O., Shah, A. P., & Robbins, T. W. (2008). Stop-signal reaction-time task performance: Role of prefrontal cortex and subthalamic nucleus. *Cereb. Cortex*, *18*, 178–188.

Eblen, F., & Graybiel, A. M. (1995). Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. *J. Neurosci.*, *15*, 5999–6013.

Feger, J., Bevan, M., & Crossman, A. R. (1994). The projections from the parafascicular thalamic nucleus to the subthalamic nucleus and the striatum arise from separate neuronal populations: A comparison with the corticostriatal and corticosubthalamic efferents in a retrograde fluorescent double-labeling study. *Neuroscience*, *60*, 125–132.

Ferry, A. T., Ongur, D., An, X., & Price, J. L. (2000). Prefrontal cortical projections to the striatum in macaque monkeys: Evidence for an organization related to prefrontal networks. *J. Comp. Neurol.*, *425*, 447–470.

Filion, M., Tremblay, L., & Bédard, P. J. (1988). Abnormal influences of passive limb movement on the activity of globus pallidus neurons in parkinsonian monkeys. *Brain Res.*, *444*, 165–176.

Filion, M., Tremblay, L., & Bedard, P. J. (1989). Excessive and unselective responses of medial pallidal neruons to both passive movement and striatal stimulation in monkeys with MPTP-induced parkinsonism. In A. R. Crossman & M. A. Sambrook (Eds.), *Neural mechanisms in disorders of movement* (pp. 157–164). London: John Libbey.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*, 1898–1902.

Flaherty, A. W., & Graybiel, A. M. (1991). Corticostriatal transformations in the primate somatosensory system: Projections from physiologically mapped body-part representations. *J. Neurophysiol.*, *66*, 1249–1263.

Flaherty, A. W., & Graybiel, A. M. (1994). Input-output organization of the sensorimotor striatum in the squirrel monkey. *J. Neurosci.*, *14*, 599–610.

Francois, C., Jan, C., McCairn, K., Grabli, D., Hirsch, E. C., Feger, J., & Tremblay, L. (2002). A primate model of Tourette's syndrome: Anatomical analysis of the basal ganglia territories involved in hyperactivity disorder with attention deficit (HD/AD) and stereotypy. *Soc. Neurosci. Abstr.*, *23*, 663.6.

Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Net.*, *19*, 1120–1136.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001a). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cogn. Affect. Behav. Neurosci.*, *1*, 137–160.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001b). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognit. Affect. Behav. Neurosci.*, *1*, 137–160.

Fudge, J. L., Kunishio, K., Walsh, P., Richard, C., & Haber, S. N. (2002). Amygdaloid projections to ventromedial striatal subterritories in the primate. *Neuroscience*, *110*, 257–275.

Fujii, N., & Graybiel, A. (2003). Representation of action sequence boundaries by macaque prefrontal cortical neurons. *Science*, *301*, 1246–1249.

Fuster, J. M. (1997). *The prefrontal cortex: Anatomy, physiology, and neuropsychology of the frontal lobe* (3rd ed.). Philadelphia: Lippincott-Raven.

Georgiou, N., Bradshaw, J. L., Phillips, J. G., & Chiu, E. (1996). The effect of Huntington's disease and Gilles de la Tourette's syndrome on the ability to hold and shift attention. *Neuropsychologia*, *34*, 843–851.

Gillies, A., & Arbuthnott, G. (2000). Computational models of the basal ganglia. *Mov. Disord.*, *15*, 762–770.

Glimcher, P. W. (2003). The neurobiology of visual-saccadic decision making. *Annu. Rev. Neurosci.*, *26*, 133–179.

Goldberg, J. A., Boraud, T., Maraton, S., Haber, S. N., Vaadia, E., & Bergman, H. (2002). Enhanced synchrony among primary motor cortex neurons in the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine primate model of Parkinson's disease. *J. Neurosci.*, *22*, 4639–4653.

Goldman, P. S., & Rosvold, H. E. (1972). The effects of selective caudate lesions in infant and juvenile rhesus monkeys. *Brain Res.*, *43*, 53–66.

Grafton, S., Hazeltine, E., & Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *J. Cogn. Neurosci.*, *7*, 497–510.

Graybiel, A. M. (1978). Organization of the nigrotectal connection: An experimental tracer study in the cat. *Brain Res.*, *143*, 339–348.

Graybiel, A. M. (1984). Correspondence between the dopamine islands and striosomes of the mammalian striatum. *Neuroscience*, *13*, 1157–1187.

Graybiel, A. M. (1995). Building action repertoires: Memory and learning functions of the basal ganglia. *Curr. Opin. Neurobiol.*, *5*, 733–741.

Graybiel, A. M. (1997). The basal ganglia and cognitive pattern generators. *Schizophr. Bull.*, *23*, 459–469.

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.*, *70*, 119–136.

Graybiel, A. M. (2005). The basal ganglia: Learning new tricks and loving it. *Curr. Opin. Neurobiol.*, *15*, 638–644.

Graybiel, A. M. (2008). Habits, rituals and the evaluative brain. *Annu. Rev. Neurosci.*, *31*, 359–387.

Graybiel, A. M., Aosaki, T., Flaherty, A. W., & Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science*, *265*, 1826–1831.

Graybiel, A. M., & Canales, J. J. (2001). The neurobiology of repetitive behaviors: Clues to the neurobiology of Tourette syndrome. In D. Cohen, C. Goetz, & J. Jankovic (Eds.), *Tourette syndrome* (pp. 123–131). Philadelphia: Williams & Wilkins.

Graybiel, A. M., Canales, J. J., & Capper-Loup, C. (2000). Levodopa-induced dyskinesias and dopamine-dependent stereotypies: A new hypothesis. *Trends Neurosci.*, *23*, S71–S77.

Graybiel, A. M., & Kubota, Y. (2003). Understanding basal ganglia function as part of a habit formation system. In M.-A. Bédard, Y. Agid, S. Chouinard, S. Fahn, A. Korczyn, & P. Lesperance (Eds.), *Mental and behavioral dysfunction in movement disorders* (pp. 51–57). Totowa, NJ: Humana.

Graybiel, A. M., & Ragsdale, C. W., Jr. (1978). Histochemically distinct compartments in the striatum of human, monkey, and cat demonstrated by acetylthiocholinesterase staining. *Proc. Natl. Acad. Sci. USA*, *75*, 5723–5726.

Graybiel, A. M., & Ragsdale, C. W., Jr. (1979). Fiber connections of the basal ganglia. In M. Cuénod, G. W. Kreutzberg, &

F. E. Bloom (Eds.), *Development and chemical specificity of neurons* (pp. 239–283). Amsterdam: Elsevier.

Graybiel, A. M., & Rauch, S. L. (2000). Toward a neurobiology of obsessive-compulsive disorder. *Neuron*, *28*, 343–347.

Groenewegen, H. J., Wright, C. I., & Uylings, H. B. (1997). The anatomical relationships of the prefrontal cortex with limbic structures and the basal ganglia. *J. Psychopharmacol.*, *11*, 99–106.

Haber, S., & McFarland, N. R. (2001). The place of the thalamus in frontal cortical-basal ganglia circuits. *Neuroscientist*, *7*, 315–324.

Haber, S. N., Kim, K. S., Mailly, P., & Calzavara, R. (2006). Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *J. Neurosci.*, *26*, 8368–8376.

Hazeltine, E., Grafton, S. T., & Ivry, R. (1997). Attention and stimulus characteristics determine the locus of motor-sequence encoding: A PET study. *Brain 120*(Pt. 1), 123–140.

Heindel, W. C., Butters, N., & Salmon, D. P. (1988). Impaired learning of a motor skill in patients with Huntington's disease. *Behav. Neurosci.*, *102*, 141–147.

Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., & Doya, K. (1999). Parallel neural networks for learning sequential procedures. *Trends Neurosci.*, *22*, 464–471.

Hikosaka, O., & Wurtz, R. H. (1985). Modification of saccadic eye movements by GABA-related substances: II. Effects of mus in monkey substantia nigra pars reticulata. *J. Neurophysiol.*, *53*, 292–308.

Honda, M., Deiber, M. P., Ibanez, V., Pascual-Leone, A., Zhuang, P., & Hallett, M. (1998). Dynamic cortical involvement in implicit and explicit motor sequence learning: A PET study. *Brain*, *121*, 2159–2173.

Ito, R., Dalley, J. W., Robbins, T. W., & Everitt, B. J. (2002). Dopamine release in the dorsal striatum during cocaine-seeking behavior under the control of a drug-associated cue. *J. Neurosci.*, *22*, 6247–6253.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Comput.*, *3*, 79–87.

James, W. (1890). *The principles of psychology* (1950 ed.). New York: Dover.

Jenkins, I. H., Brooks, D. J., Nixon, P. D., Frackowiak, R. S., & Passingham, R. E. (1994). Motor sequence learning: A study with positron emission tomography. *J. Neurosci.*, *14*, 3775–3790.

Joel, D. (2001). Open interconnected model of basal ganglia-thalamocortical circuitry and its relevance to the clinical syndrome of Huntington's disease. *Mov. Disord.*, *16*, 407–423.

Jog, M., Kubota, Y., Connolly, C. I., Hillegaart, V., & Graybiel, A. M. (1999). Building neural representations of habits. *Science*, *286*, 1745–1749.

Jueptner, M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. J., & Passingham, R. E. (1997). Anatomy of motor learning: II. Subcortical structures and learning by trial and error. *J. Neurophysiol.*, *77*, 1325–1337.

Jueptner, M., Stephan, K. M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. J., & Passingham, R. E. (1997). Anatomy of motor learning: I. Frontal cortex and attention to action. *J. Neurophysiol.*, *77*, 1313–1324.

Jueptner, M., & Weiller, C. (1998). A review of differences between basal ganglia and cerebellar control of movements as revealed by functional imaging studies. *Brain*, *121*, 1437–1449.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M., Turner, R., et al. (1998). The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proc. Natl. Acad. Sci. USA*, *95*, 861–868.

Kermadi, I., & Boussaoud, D. (1995). Role of the primate striatum in attention and sensorimotor processes: Comparison with premotor cortex. *NeuroReport*, *6*, 1177–1181.

Kincaid, A. E., Zheng, T., & Wilson, C. J. (1998). Connectivity and convergence of single corticostriatal axons. *J. Neurosci.*, *18*, 4722–4731.

Kitano, K., Aoyagi, T., & Fukai, T. (2001). A possible functional organization of the corticostriatal input within the weakly-correlated striatal activity: A modeling study. *Neurosci. Res.*, *40*, 87–96.

Klein, D., Zatorre, R. J., Milner, B., Meyer, E., & Evans, A. C. (1994). Left putaminal activation when speaking a second language: Evidence from PET. *NeuroReport*, *5*, 2295–2297.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learn. Mem.*, *1*, 106–120.

Kolomiets, B. P., Deniau, J. M., Mailly, P., Menetrey, A., Glowinski, J., & Thierry, A. M. (2001). Segregation and convergence of information flow through the cortico-subthalamic pathways. *J. Neurosci.*, *21*, 5764–5772.

Koski, L., Paus, T., Hofle, N., & Petrides, M. (1999). Increased blood flow in the basal ganglia when using cues to direct attention. *Exp. Brain Res.*, *129*, 241–246.

Krack, P., Kumar, R., Ardouin, C., Dowsey, P. L., McVicker, J. M., Benabid, A. L., et al. (2001). Mirthful laughter induced by subthalamic nucleus stimulation. *Mov. Disord.*, *16*, 867–875.

Lang, A. E. (2000). Surgery for levodopa-induced dyskinesias. *Ann. Neurol.*, *47*, S193–S199; discussion S199–S202.

Lansink, C. S., Goltstein, P. M., Lankelma, J. V., Joosten, R. N., McNaughton, B. L., & Pennartz, C. M. (2008). Preferential reactivation of motivationally relevant information in the ventral striatum. *J. Neurosci.*, *28*, 6372–6382.

Lavoie, B., & Parent, A. (1994). Pedunculopontine nucleus in the squirrel monkey: Distribution of cholinergic and monoaminergic neurons in the mesopontine tegmentum with evidence for the presence of glutamate in cholinergic neurons. *J. Comp. Neurol.*, *344*, 190–209.

Lazaro, L., Caldu, X., Junque, C., Bargallo, N., Andres, S., Morer, A., et al. (2008). Cerebral activation in children and adolescents with obsessive-compulsive disorder before and after treatment: A functional MRI study. *J. Psychiatr. Res.*, *42*, 1051–1059.

Leckman, J. F. (2002). Tourette's syndrome. *Lancet*, *360*, 1577–1586.

Lei, W., Jiao, Y., Del Mar, N., & Reiner, A. (2004). Evidence for differential cortical input to direct pathway versus indirect pathway striatal projection neurons in rats. *J. Neurosci.*, *24*, 8289–8299.

Leichnetz, G. R. (2001). Connections of the medial posterior parietal cortex (area 7m) in the monkey. *Anat. Rec.*, *263*, 215–236.

Malach, R., & Graybiel, A. M. (1986). Mosaic architecture of the somatic sensory-recipient sector of the cat's striatum. *J. Neurosci.*, *6*, 3436–3458.

Martone, M., Butters, N., Payne, M., Becker, J. T., & Sax, D. S. (1984). Dissociations between skill learning and verbal recognition in amnesia and dementia. *Arch. Neurol.*, *41*, 965–970.

Matsumoto, N., Minamimoto, T., Graybiel, A. M., & Kimura, M. (2001). Neurons in the thalamic CM-Pf complex supply neurons in the striatum with information about behaviorally significant sensory events. *J. Neurophysiol.*, *85*, 960–976.

McAlonan, K., & Brown, V. J. (2002). The thalamic reticular nucleus: More than a sensory nucleus? *Neuroscientist*, *8*, 302–305.

McDonald, R. J., & White, N. M. (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behav. Neurosci.*, *107*, 3–22.

McDonald, R. J., & White, N. M. (1994). Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. *Behav. Neural Biol.*, *61*, 260–270.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nat. Neurosci.*, *11*, 103–107.

Mendez, M. F., Adams, N. L., & Lewandowski, K. S. (1989). Neurobehavioral changes associated with caudate lesions. *Neurology*, *39*, 349–354.

Mesulam, M.-M. (2000). Attention, confusional states and neglect. In M.-M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (2nd ed., pp. 125–168). New York: Oxford University Press.

Middleton, F. A., & Strick, P. L. (2002). Basal-ganglia 'projections' to the prefrontal cortex of the primate. *Cereb. Cortex*, *12*, 926–935.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Prog. Neurobiol.*, *50*, 381–425.

Mink, J. W. (2001). Basal ganglia dysfunction in Tourette's syndrome: A new hypothesis. *Pediatr. Neurol.*, *25*, 190–198.

Mink, J. W. (2003). The basal ganglia. In L. R. Squire, F. E. Bloom, S. K. McConnell, J. L. Roberts, N. C. Spitzer, & M. J. Zigmond (Eds.), *Fundamental neuroscience* (pp. 815–839). San Diego, CA: Academic.

Mink, J. W., & Thach, W. T. (1991). Basal ganglia motor control: III. Pallidal ablation: Normal reaction time, muscle cocontraction, and slow movement. *J. Neurophysiol.*, *65*, 330–351.

Miyachi, S., Hikosaka, O., Miyashita, K., Karadi, Z., & Rand, M. K. (1997). Differential roles of monkey striatum in learning of sequential hand movement. *Exp. Brain Res.*, *115*, 1–5.

Miyashita, N., Hikosaka, O., & Kato, M. (1995). Visual hemineglect induced by unilateral striatal dopamine deficiency in monkeys. *NeuroReport*, *6*, 1257–1260.

Monakow, K. H., Akert, K., & Kunzle, H. (1978). Projections of the precentral motor cortex and other cortical areas of the frontal lobe to the subthalamic nucleus in the monkey. *Exp. Brain Res.*, *33*, 395–403.

Nakamura, K., Sakai, K., & Hikosaka, O. (1998). Neuronal activity in medial frontal cortex during learning of sequential procedures. *J. Neurophysiol.*, *80*, 2671–2687.

Nakamura, K., Sakai, K., & Hikosaka, O. (1999). Effects of local inactivation of monkey medial frontal cortex in learning of sequential procedures. *J. Neurophysiol.*, *82*, 1063–1068.

Nambu, A., Tokuno, H., & Takada, M. (2002). Functional significance of the cortico-subthalamo-pallidal "hyperdirect" pathway. *Neurosci. Res.*, *43*, 111–117.

Nambu, A., Takada, M., Inase, M., & Tokuno, H. (1996). Dual somatotopical representations in the primate subthalamic nucleus: Evidence for ordered but reversed body-map transformations from the primary motor cortex and the supplementary motor area. *J. Neurosci.*, *16*, 2671–2683.

Nambu, A., Tokuno, H., Hamada, I., Kita, H., Imanishi, M., Akazawa, T., et al. (2000). Excitatory cortical inputs to pallidal neurons via the subthalamic nucleus in the monkey. *J. Neurophysiol.*, *84*, 289–300.

Nambu, A., Yoshida, S., & Jinnai, K. (1990). Discharge patterns of pallidal neurons from various cortical areas during movement in the monkey. *Brain Res.*, *519*, 183–191.

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cogn. Affect. Behav. Neurosci.*, *3*, 17–26.

Niv, Y., Duff, M. O., & Dayan, P. (2005). Dopamine, uncertainty and TD learning. *Behav. Brain Funct.*, *1*, 6–14.

Niv, Y., Joel, D., & Dayan, P. (2006). A normative perspective on motivation. *Trends Cogn. Sci.*, *10*, 375–381.

Oberg, R. G. E., & Divac, I. (1979). "Cognitive" functions of the neostriatum. In I. Divac & R. G. E. Öberg (Eds.), *The neostriatum* (pp. 291–313). London: Pergamon.

Obeso, J. A., Rodriguez-Oroz, M. C., Rodriguez, M., Macias, R., Alvarez, L., Guridi, J., Vitek, J., & DeLong, M. R. (2000). Pathophysiologic basis of surgery for Parkinson's disease. *Neurology*, *55*, S7–S12.

Owen, A. M., Roberts, A. C., Hodges, J. R., Summers, B. A., Polkey, C. E., & Robbins, T. W. (1993). Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease. *Brain*, *116*(Pt. 5), 1159–1175.

Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *J. Neurosci.*, *9*, 1465–1472.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.*, *25*, 563–593.

Packard, M. G., & McGaugh, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behav. Neurosci.*, *106*, 439–446.

Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol. Learn. Mem.*, *65*, 65–72.

Pahapill, P. A., & Lozano, A. M. (2000). The pedunculopontine nucleus and Parkinson's disease. *Brain*, *123*(Pt. 9), 1767–1783.

Parent, A., & Hazrati, L. -N. (1993). Anatomical aspects of information processing in primate basal ganglia. *Trends. Neurosci.*, *16*, 111–116.

Parent, A., & Hazrati, L. N. (1995a). Functional anatomy of the basal ganglia: II. The place of subthalamic nucleus and external pallidum in basal ganglia circuitry. *Brain Res. Brain Res. Rev.*, *20*, 128–154.

Parent, A., & Hazrati, L. N. (1995b). Functional anatomy of the basal ganglia: I. The corticobasal ganglia-thalamo-cortical loop. *Brain Res. Brain Res. Rev.*, *20*, 91–127.

Parent, A., Mackey, A., & De Bellefeuille, L. (1983). The subcortical afferents to caudate nucleus and putamen in primate: A fluorescence retrograde double labeling study. *Neuroscience*, *10*, 1137–1150.

Parthasarathy, H. B., & Graybiel, A. M. (1997). Cortically driven immediate-early gene expression reflects modular influence of sensorimotor cortex on identified striatal neurons in the squirrel monkey. *J. Neurosci.*, *17*, 2477–2491.

Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*, *433*, 873–876.

Paus, T. (2001). Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nat. Rev. Neurosci.*, *2*, 417–424.

Peigneux, P., Maquet, P., Meulemans, T., Destrebecqz, A., Laureys, S., Degueldre, C., Delfiore, G., Aerts, J., Luxen, A., Franck, G., Van der Linden, M., & Cleeremans, A. (2000). Striatum forever, despite sequence learning variability: A random effect analysis of PET data. *Hum. Brain Mapping*, *10*, 179–194.

Peterson, B. S., Skudlarski, P., Anderson, A. W., Zhang, H., Gatenby, J. C., Lacadie, C. M., Leckman, J. F., & Gore, J. C. (1998). A functional magnetic resonance imaging study of tic suppression in Tourette syndrome. *Arch. Gen. Psychiatry*, *55*, 326–333.

Phillips, P. E., Stuber, G. D., Heien, M. L., Wightman, R. M., & Carelli, R. M. (2003). Subsecond dopamine release promotes cocaine seeking. *Nature*, *422*, 614–618.

Pillon, B., Deweer, B., Michon, A., Malapani, C., Agid, Y., & Dubois, B. (1994). Are explicit memory disorders of progressive supranuclear palsy related to damage to striatofrontal circuits? Comparison with Alzheimer's, Parkinson's, and Huntington's diseases. *Neurology*, *44*, 1264–1270.

Poldrack, R. A., & Gabrieli, J. D. E. (2001). Characterizing the neural mechanisms of skill learning and repetition priming. *Brain*, *124*, 67–82.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology*, *13*, 564–574.

Pontone, G., Williams, J. R., Bassett, S. S., & Marsh, L. (2006). Clinical features associated with impulse control disorders in Parkinson disease. *Neurology*, *67*, 1258–1261.

Prensa, L., & Parent, A. (2001). The nigrostriatal pathway in the rat: A single-axon study of the relationship between dorsal and ventral tier nigral neurons and the striosome/matrix striatal compartments. *J. Neurosci.*, *21*, 7247–7260.

Ragsdale, C. W., Jr., & Graybiel, A. M. (1991). Compartmental organization of the thalamostriatal connection in the cat. *J. Comp. Neurol.*, *311*, 134–167.

Rao, S. M., Bobholz, J. A., Hammeke, T. A., Rosen, A. C., Woodley, S. J., Cunningham, J. M., et al. (1997). Functional MRI evidence for subcortical participation in conceptual reasoning skills. *NeuroReport*, *8*, 1987–1993.

Rauch, S. L., Whalen, P. J., Curran, T., Shin, L. M., Coffey, B. J., Savage, C. R., McInerney, S. C., Baer, L., & Jenike, M. A. (2001). Probing striato-thalamic function in obsessive-compulsive disorder and Tourette syndrome using neuroimaging methods. *Adv. Neurol.*, *85*, 207–224.

Raz, A., Feingold, A., Zelanskaya, V., Vaadia, E., & Bergman, H. (1996). Neuronal synchronization of tonically active neurons in the striatum of normal and parkinsonian primates. *J. Neurophysiol.*, *76*, 2083–2088.

Raz, A., Vaadia, E., & Bergman, H. (2000). Firing patterns and correlations of spontaneous discharge of pallidal neurons in the normal and the tremulous 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine vervet model of parkinsonism. *J. Neurosci.*, *20*, 8559–8571.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience*, *89*, 1009–1023.

Reynolds, J. N., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. *Nature*, *413*, 67–70.

Rinvik, E., Grofova, I., & Ottersen, O. P. (1976). Demonstration of nigrotectal and nigroreticular projections in the cat by axonal transport of proteins. *Brain Res.*, *112*, 388–394.

Robbins, T. W., & Everitt, B. J. (1992). Functions of dopamine in the dorsal and ventral striatum. *Semin. Neurosci.*, *4*, 119–127.

Rolls, E. T. (Ed.). (1999). *The brain and emotion.* New York: Oxford University Press.

Rowe, J., Stephan, K. E., Friston, K., Frackowiak, R., Lees, A., & Passingham, R. (2002). Attention to action in Parkinson's disease: Impaired effective connectivity among frontal cortical regions. *Brain*, *125*, 276–289.

Saint-Cyr, J. A. (2003). Frontal-striatal circuit functions: Context, sequence, and consequence. *J. Int. Neuropsychol. Soc.*, *9*, 103–127.

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostriatal dysfunction in man. *Brain*, *111*, 941–959.

Saka, E., Goodrich, C., Harlan, P., Madras, B. K., & Graybiel, A. M. (2004). Repetitive behaviors in monkeys are linked to specific striatal activation patterns. *J. Neurosci.*, *24*, 7557–7565.

Saka, E., & Graybiel, A. (2004). Pathophysiology of Tourette syndrome: Striatal pathways revisited. *Brain Dev.*, *25*(Suppl. 1), S15–S19.

Samuel, M., Ceballos-Baumann, A. O., Turjanski, N., Boecker, H., Gorospe, A., Linazasoro, G., Holmes, A. P., DeLong, M. R., Vitek, J. L., Thomas, D. G., Quinn, N. P., Obeso, J. A., & Brooks, D. J. (1997). Pallidotomy in Parkinson's disease increases supplementary motor area and prefrontal activation during performance of volitional movements: An H2(15)O PET study. *Brain*, *120*(Pt. 8), 1301–1313.

Sato, K., Sumi-Ichinose, C., Kaji, R., Ikemoto, K., Nomura, T., Nagatsu, I., Nagahiro, S., Graybiel, A. M., & Goto, S. (2008). Differential involvement of striosome-matrix dopamine systems in transgenic model for dopa-responsive dystonia: A predominant loss of striosomal dopaminergic inputs. *Proc. Natl. Acad. Sci. USA*, *105*, 12551–12556.

Sawamoto, N., Honda, M., Hanakawa, T., Aso, T., Inoue, M., Toyoda, H., Ishizu, K., Fukuyama, H., & Shibasaki, H. (2007). Cognitive slowing in Parkinson disease is accompanied by hypofunctioning of the striatum. *Neurology*, *68*, 1062–1068.

Schneider, J. S., & Lidsky, T. I. (1981). Processing of somatosensory information in the striatum of behaving cats. *J. Neurophysiol.*, *45*, 841–851.

Schroeder, J. P., Wingard, J. C., & Packard, M. G. (2002). Post-training reversible inactivation of hippocampus reveals interference between memory systems. *Hippocampus*, *12*, 280–284.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*, 241–263.

Schultz, W. (2007). Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.*, *30*, 259–288.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Schupbach, W. M., & Agid, Y. (2008). Psychosocial adjustment after deep brain stimulation in Parkinson's disease. *Nat. Clin. Pract. Neurol.*, *4*, 58–59.

Schwartz, J. M., Stoessel, P. W., Baxter, L. R., Jr., Martin, K. M., & Phelps, M. E. (1996). Systematic changes in cerebral

glucose metabolic rate after successful behavior modification treatment of obsessive-compulsive disorder. *Arch. Gen. Psychiatry*, *53*, 109–113.

Shadmehr, R., & Brashers-Krug, T. (1997). Functional stages in the formation of human long-term motor memory. *J. Neurosci.*, *17*, 409–419.

Shidara, M., Aigner, T. G., & Richmond, B. J. (1998). Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J. Neurosci.*, *18*, 2613–2625.

Sieradzan, K. A., & Mann, D. M. (2001). The selective vulnerability of nerve cells in Huntington's disease. *Neuropathol. Appl. Neurobiol.*, *27*, 1–21.

Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C., & Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolate: From pleasure to aversion. *Brain*, *124*, 1720–1733.

Smith, A. D., & Bolam, J. P. (1990). The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurons. *Trends Neurosci.*, *13*, 259–265.

Smith, M. A., Brandt, J., & Shadmehr, R. (2000). Motor disorder in Huntington's disease begins as a dysfunction in error feedback control. *Nature*, *403*, 544–549.

Smith, Y., Bevan, M. D., Shink, E., & Bolam, J. P. (1998). Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience*, *86*, 353–387.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA: MIT Press.

Swerdlow, N. R., & Koob, G. F. (1987). Dopamine, schizophrenia, mania, and depression: Toward a unified hypothesis of cortico-striato-pallido-thalamic function. *Behav. Brain Res.*, *10*, 197–245.

Taha, S. A., Nicola, S. M., & Fields, H. L. (2007). Cue-evoked encoding of movement planning and execution in the rat nucleus accumbens. *J. Physiol.*, *584*, 801–818.

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., & Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat. Neurosci.*, *7*, 887–893.

Tang, C., Pawlak, A. P., Prokopenko, V., & West, M. O. (2007). Changes in activity of the striatum during formation of a motor habit. *Eur. J. Neurosci.*, *25*, 1212–1227.

Teicher, M. H., Anderson, C. M., Polcari, A., Glod, C. A., Maas, L. C., & Renshaw, P. F. (2000). Functional deficits in basal ganglia of children with attention-deficit/hyperactivity disorder shown with functional magnetic resonance imaging relaxometry. *Nat. Med.*, *6*, 470–473.

Tremblay, L., Grabli, D., McCairn, K., Jan, C., Hirsch, E., Feger, J., & Francois, C. (2003). A monkey model of Tourette's syndrome: Induction of hyperactivity disorder with attention deficit (HD/AD) and stereotypy by microinjections of bicuculline in the external segment of globus pallidus. *Soc. Neurosci. Abstr.*, *23*, 663.7.

Turner, R. S., & DeLong, M. R. (2000). Corticostriatal activity in primary motor cortex of the macaque. *J. Neurosci.*, *20*, 7096–7108.

Ungerleider, L. G., Doyon, J., & Karni, A. (2002). Imaging brain plasticity during motor skill learning. *Neurobiol. Learn Mem.*, *78*, 553–564.

van den Heuvel, O. A., Groenewegen, H. J., Barkhof, F., Lazeron, R. H. C., van Dyck, R., & Veltman, D. J. (2003). Frontostriatal system in planning complexity: A parametric functional magnetic resonance version of Tower of London task. *NeuroImage*, *18*, 367–374.

Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1993). Subcortical connections of inferior temporal areas TE and TEO in macaque monkeys. *J. Comp. Neurol.*, *335*, 73–91.

Weintraub, D., Siderowf, A. D., Potenza, M. N., Goveas, J., Morales, K. H., Duda, J. E., Moberg, P. J., & Stern, M. B. (2006). Association of dopamine agonist use with impulse control disorders in Parkinson disease. *Arch. Neurol.*, *63*, 969–973.

White, N. M., & Hiroi, N. (1998). Preferential localization of self-stimulation sites in striosomes/patches in the rat striatum. *Proc. Natl. Acad. Sci. USA*, *95*, 6486–6491.

Willingham, D. B., Salidis, J., & Gabrieli, J. D. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *J. Neurophysiol.*, *88*, 1451–1460.

Winstanley, C. A., Eagle, D. M., & Robbins, T. W. (2006). Behavioral models of impulsivity in relation to ADHD: Translation between clinical and preclinical studies. *Clin. Psychol. Rev.*, *26*, 379–395.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nat. Rev. Neurosci.*, *5*, 483–494.

Yeterian, E. H., & Pandya, D. N. (1998). Corticostriatal connections of the superior temporal region in rhesus monkeys. *J. Comp. Neurol.*, *399*, 384–402.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.*, *7*, 464–476.

Young, A. B., Albin, R. L., & Penney, J. B. (1989). Neuropharmacology of basal ganglia functions: Relationship to pathophysiology of movement disorders. In A. R. Crossman & M. A. Sambrook (Eds.), *Neural mechanisms in disorders of movement. Current problems in neurology* (Vol. 9, pp. 17–27). London: John Libbey.

Zald, D. H., Boileau, I., El-Dearedy, W., Gunn, R., McGlone, F., Dichter, G. S., et al. (2004). Dopamine transmission in the human striatum during monetary reward tasks. *J. Neurosci.*, *24*, 4105–4112.

40 Computational Neuroanatomy of Voluntary Motor Control

REZA SHADMEHR AND JOHN W. KRAKAUER

ABSTRACT We review some of the impairments in motor control, motor learning, and higher-order motor control in patients with lesions of the cerebellum, parietal cortex, and basal ganglia. We attempt to explain some of these impairments in terms of computational ideas such as state estimation, optimization, prediction, cost, and reward. We suggest that a function of the cerebellum is system identification: to built internal models that predict sensory outcome of motor commands and correct motor commands through internal feedback. A function of the parietal cortex is state estimation: to integrate the predicted proprioceptive and visual outcomes with sensory feedback to form a belief about how the commands affected the states of the body and the environment. A function of basal ganglia is related to optimal control: learning costs and rewards associated with sensory states and estimating the "cost-to-go" during execution of a motor task.

Over the last 25 years, a large body of experimental and theoretical work has been directed toward understanding the computational basis of motor control, particularly visually guided reaching. Roboticists and engineers largely initiated this work, with the aim of deriving from first principles some of the strikingly stereotypical features of movements observed in people and other primates. That is, they aimed to understand why we move the way that we do. The theories began to explain why in reaching to pick up a cup or in moving the eyes to look at an object, there was such consistency in the detailed trajectory of the hand and the eyes. In many ways, the approach was reminiscent of physics and its earliest attempts to explain regularity in motion of celestial objects except that the regularity was in our movements, and the search was for theories that explained our behavior. Here, we will summarize these theories and then link them to experimental findings in healthy subjects and in patients with neurological disease.

REZA SHADMEHR Laboratory for Computational Motor Control, Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, Maryland

JOHN W. KRAKAUER The Motor Performance Laboratory, Department of Neurology, Columbia University College of Physicians and Surgeons, New York, New York

The computational problem of motor control

In 1954, Fitts published a short paper in which he reported that there were regularities in people's movements (Fitts, 1954). He asked volunteers to move a pen from one "goal region" to another as fast and accurately as they could. He found that the movement durations grew logarithmically as a function of the distance between the goals (figure 40.1). This relationship was modulated by two factors. One factor was the size of the goal region. As the goal region became smaller, movements slowed down. A second factor was the mass of the pen. People slowed their movements when they moved a heavier pen. To explain these results, consider that the target box was surrounded by two penalty regions, so it seems rational to aim for the center of the target box. What if the penalty region was only on one side? Now one should aim for a point farther away from the penalty region and not at the center of the target box (Trommershauser, Gepshtein, Maloney, Landy, & Banks, 2005). This is because movements have variability, and one will maximize reward (in terms of sum of hits and misses) if one takes into account this variability. This variability explains the speed of movements in Fitt's experiment and the sensitivity to pen weight: Rapid movements are more variable than slow movements, so one should slow down if there is a need to be accurate. Moving heavier objects tends to increase movement variability, again requiring a reduced speed to maintain accuracy. Therefore in planning our movements, our brain takes into account movement variability because variability affects accuracy, which in turn affects our ability to acquire reward.

Harris and Wolpert (1998) began formalizing these ideas by linking variability and movement planning. They noted that larger motor commands required larger neural activity, which in turn produced larger variability owing to a noise process that grew with the mean of the signal. Therefore, motor commands carried an accuracy cost because the larger the command, the larger the standard deviation of the noise that rides on top of the force produced by the muscles (Jones, Hamilton, & Wolpert, 2002). Noise makes movements inaccurate.

In a sense, the theory restated the purpose of movements using language of mathematics: Be as fast as possible,

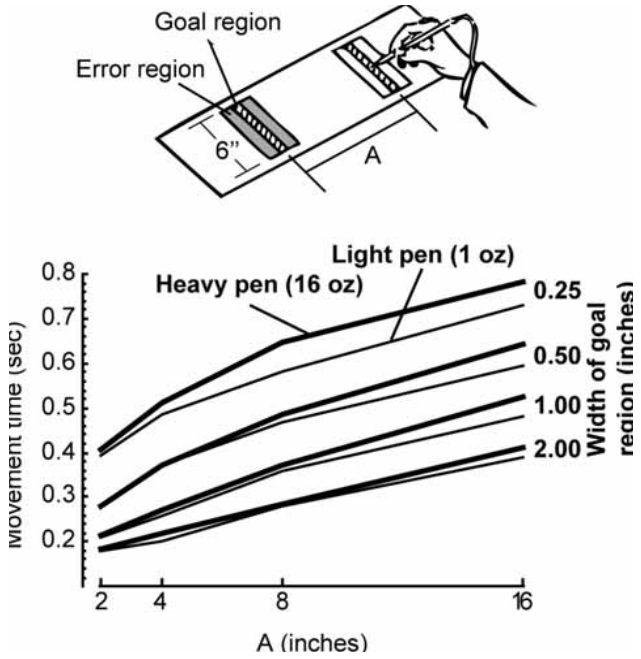


FIGURE 40.1 Accuracy constraints affect control of reaching. Volunteers were instructed to tap the two goal regions with a pen as many times as possible during a 15-s period. Movement time increased as the accuracy requirements increased (width of target region decreased) and as the weight of the hand-held pen increased. (Figure constructed from data in Fitts, 1954.)

while trying to be as accurate as the requirements imposed by the task. However, by doing so, it forced the theorists to think how one would actually achieve this optimality. Certainly, the solution to the problem could not be "hard wired."

First, costs and rewards of tasks are not constant. Take the simple saccade task in which an animal is given more reward for certain visual targets and less for others. Hikosaka and colleagues (Takikawa, Kawagoe, Itoh, Nakahara, & Hikosaka, 2002) examined eye trajectories when a monkey was asked to make saccades to various target locations. They noted that peak speeds tended to be higher and less variable when saccades were to rewarded target locations. Therefore, when the expected rewards of the task change, movement planning responds to these changes.

Second, the brain alters movement planning as the dynamics of the body or a tool change (e.g., the light versus heavy pens in figure 40.1). That is, the nervous system cannot rely on a motor plant that is time-invariant. Rather, it seems more reasonable that the nervous system should monitor these changes and form an *internal model* of the plant and/or the tool (Shadmehr & Mussa-Ivaldi, 1994). Indeed, maintaining performance in something as simple as a saccade or a reach probably requires constant adjustment of this internal model (Smith, Ghazizadeh, & Shadmehr, 2006; Kording, Tenenbaum, & Shadmehr, 2007).

Todorov and Jordan (2002) recognized that a key component of the problem was presence of feedback. One type of feedback is from sensory receptors that monitor the state of the body and the world. Another type of feedback is from internal models that monitor the motor output and predict their sensory consequences, effectively providing a form of internal feedback. Internal predictions can be made long before sensory feedback, making some very rapid movements such as saccades depend entirely on internal feedback (Chen-Harris, Joiner, Ethier, Zee, & Shadmehr, 2008). However, for longer movements, the two kinds of information would need to be combined to form a belief about the state of the body. Todorov and Jordan (2002) suggested that a more appropriate mathematical approach was to first describe the constraints of the task in terms of a function that included explicit terms for gains and losses and then maximize that function in the framework of *feedback control*. This new formulation was a breakthrough because it formally linked motor costs, expected rewards, noise, sensory feedback, and internal models into a single, coherent mathematical framework (see chapter 42 for a thorough introduction).

We summarize this framework in figure 40.2*A*. At the heart of the approach is the idea that we make movements to achieve a rewarding state. The rewards we expect to get and the costs we expect to pay determine the trajectory we choose to execute and how we will respond to sensory feedback. To make the "best" movement, our brain needs to solve three kinds of problems: We need to be able to accurately predict the sensory consequences of our motor commands (this is called *system identification*), we need to combine these predictions with actual sensory feedback to form a belief about the state of our body and the world (called *state estimation*), and then given this belief about the state of our body and the world, we have to adjust the gains of the sensorimotor feedback loops so that our movements maximize some measure of performance (called *optimal control*).

Here, we will suggest a specific computational neuroanatomy of the motor system (figure 40.2*B*). In this framework, the basal ganglia help to form the expected costs of the motor commands and the expected rewards of the sensory states. The cerebellum plays the role of predicting the sensory consequences of motor commands, that is, the expected changes in proprioceptive and visual feedback. The parietal cortex combines the expected sensory feedback with the actual sensory feedback, computing a belief about the current proprioceptive and visual states. Given the motor costs and expected rewards of the sensory states, the premotor and the primary motor cortex assign "feedback gains" to the visual and proprioceptive states, respectively, resulting in sensorimotor maps that transform the internal belief about states into motor commands.

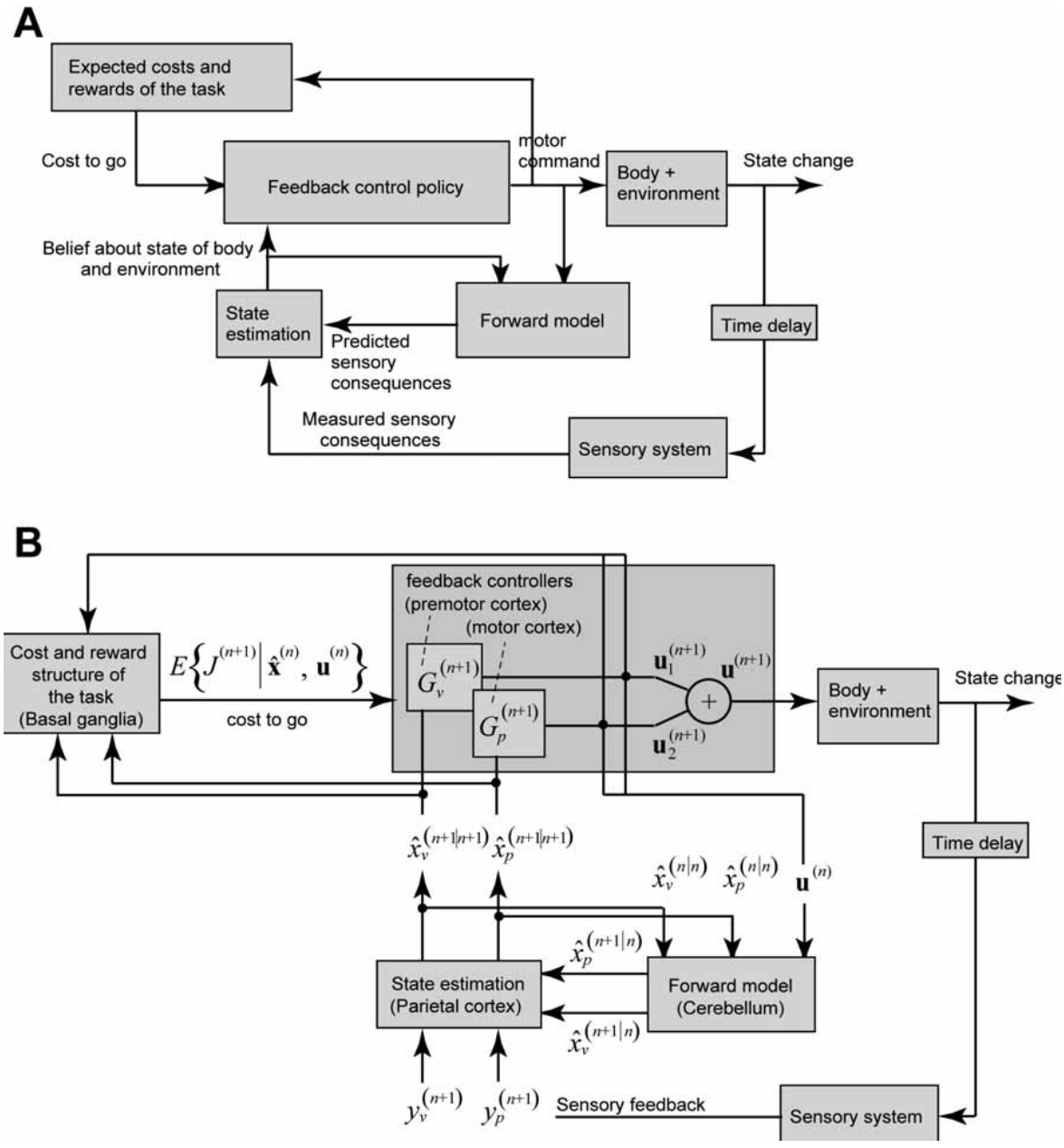


FIGURE 40.2 A schematic model for generating goal-directed movements. See the text for explanation of variables and box labels.

The computational problem in reaching

Let us use the well-studied reach adaptation paradigm to formulate the problem in the framework outlined in figure 40.2. What are the costs and rewards of a reaching task? Suppose that we are instructed to hold a tool and move it so that a cursor displayed on a monitor arrives at a target. If we accomplish this in a specific time period, we are provided a monetary reward, or juice, or perhaps a "target explosion." We can sense the position of the cursor $\mathbf{y}_v$ and the target $\mathbf{r}$ via vision and position of our arm $\mathbf{y}_p$ via proprioception. Through experience in the task, we learn that the objective is to minimize the quantity $(\mathbf{y}_v^{(t)} - \mathbf{r})^T (\mathbf{y}_v^{(t)} - \mathbf{r})$ at time $t = N$ after the reach starts (e.g., this is the time that the movement is rewarded if the cursor is in the target). Superscript T is the transpose operator. To denote the fact that this cost is zero except for time N, we write it as

$$\sum_{t=1}^{N} \left(\mathbf{y}_v^{(t)} - \mathbf{r}\right)^T Q^{(t)} \left(\mathbf{y}_v^{(t)} - \mathbf{r}\right)$$

where the matrix Q is a measure of our cost at each time step (which may be zero except at time N). That is, matrix Q specifies how important it may be for us to put the cursor

in the target. If we value the reward, then we set this variable to be large.

There is also a cost associated with motor commands **u**. This cost may reflect a desire to be as frugal as possible with our energy expenditure, or it may reflect the fact that the larger the motor commands, the larger the noise in the forces that are produced by the muscles, resulting in variability. This variability increases the difficulty in controlling the movement. As a result, we want to produce the smallest amount of motor commands possible. Now the total cost becomes

$$\mathcal{J} = \sum_{t=1}^{N} \left(\mathbf{y}_v^{(t)} - \mathbf{r}\right)^T Q^{(t)} \left(\mathbf{y}_v^{(t)} - \mathbf{r}\right) + \mathbf{u}^{(t)T} L \mathbf{u}^{(t)} \tag{1}$$

where matrix L is a measure of the costs associated with the motor commands. The relative weight of Q and L is an internal measure of expected value of achieving the goal versus expected motor costs.

To be successful in this task (consistently arrive at the target in time), we need to find the motor commands that, on the one hand, are as small as possible and, on the other hand, are large enough to get the cursor to the target in time. To do so, we need some way to relate motor commands to their outcomes. This is called an internal model. For example, through observation, we learn that moving the tool moves the cursor on the screen. In particular, motor commands $\mathbf{u}^{(t)}$ are expected to produce proprioceptive and visual feedback $\hat{\mathbf{y}}^{(t)} = [\hat{\mathbf{y}}_v^{(t)}, \hat{\mathbf{y}}_p^{(t)}]$. These are the expected sensory consequences of our action. Here, we write this "internal model" as a linear function of motor commands:

$$\begin{aligned} \hat{\mathbf{x}}^{(t+1|t)} &= \hat{A}\hat{\mathbf{x}}^{(t|t)} + \hat{B}\mathbf{u}^{(t)} \\ \hat{\mathbf{y}}^{(t)} &= \hat{H}\hat{\mathbf{x}}^{(t)} \end{aligned} \tag{2}$$

where $\hat{\mathbf{x}}^{(t|t)}$ represents the predicted state (of our body and the world) at time t given the sensory feedback up until that time, H is a transformation of those states to expected sensory feedback $\hat{\mathbf{y}}^{(t)}$ (i.e., proprioception and vision), and $\hat{\mathbf{x}}^{(t+1|t)}$ is predicted state at time $t + 1$ given the state and motor command at time t. Equation 2 describes an internal model of the dynamical system that we are trying to control. The actual dynamics of that system may be more complicated. For example, the motor commands may carry signal-dependent noise $\mathring{\mathbf{a}}_u^{(t)}$, that is, a noise in which the standard deviation grows with the size of the motor command. In general, there may be similar signal-dependent noises on our sensory system, $\mathring{\mathbf{a}}_y^{(t)}$. In sum, a reasonable representation of the stochastic system that we are trying to control might be written as

$$\begin{aligned} \mathbf{x}^{(t+1)} &= A\mathbf{x}^{(t)} + B\left(\mathbf{u}^{(t)} + \mathring{\mathbf{a}}_u^{(t)}\right) \\ \mathbf{y}^{(t)} &= H\left(\mathbf{x}^{(t)} + \mathring{\mathbf{a}}_y^{(t)}\right) \end{aligned} \tag{3}$$

In Equation 1, we introduced a motor cost $\mathbf{u}^{(t)T} L \mathbf{u}^{(t)}$, i.e., the larger the motor commands, the larger the cost. Using Equation 3, we can now give a rationale for this cost: the larger the motor commands, the larger the variance in the state of the system that we are trying to control. Therefore, motor costs implicitly attempt to reduce the variance of the movement.

As motor commands are generated, we receive a continuous stream of sensory feedback **y**. We combine the predicted sensory feedback with the observed quantities to form a *belief* about states:

$$\hat{\mathbf{x}}^{(t+1|t+1)} = \hat{\mathbf{x}}^{(t+1|t)} + K^{(t+1)}\left(\mathbf{y}^{(t+1)} - \hat{\mathbf{y}}^{(t+1)}\right) \tag{4}$$

In this equation, the term $\hat{\mathbf{x}}^{(t+1|t+1)}$ is the belief state at time $t + 1$, given that we have acquired sensory information at that time. K is a mixing gain (or a Kalman gain) that determines how much we should change our belief on the basis of the difference between what we predicted and what we observed. Therefore, equation 2 describes how we make predictions about sensory feedback, and equation 4 describes how we combine the actual sensory observations with predictions to update beliefs about states.

Our task is to perform the movement in a way that maximizes our chances for reward. If equation 2 is an accurate model of how motor commands produce changes in the states, then we can use it as a set of constraints with which to minimize equation 1. Because there is noise in our system, the cost $\mathcal{J}$ in equation 1 is a stochastic variable. At each time point during a movement, the best that we can do is minimize the expected value of this cost, given the state that we believe to be in and the motor commands that we have produced: $E\{\mathcal{J}^{(t)} | \hat{\mathbf{x}}^{(t-1)}, \mathbf{u}^{(t-1)}\}$. The term $E\{\mathcal{J}^{(t)}\}$ reflects the expected value of the cost-to-go, that is, the total cost remaining in the current trial. The result is a feedback control "gain":

$$\begin{aligned} \mathbf{u}^{(t)} &= -G^{(t)}\hat{\mathbf{x}}^{(t|t-1)} \\ &= -G_p^{(t)}\hat{\mathbf{x}}_p^{(t|t-1)} - G_v^{(t)}\hat{\mathbf{x}}_v^{(t|t-1)} \end{aligned} \tag{5}$$

The new variable G is a matrix that changes with time during a movement. It tells us how at time t, we can transform beliefs in sensory states (in terms of proprioception and vision) into motor commands so that we maximize performance in the remaining task time. Therefore, in this framework, motor planning refers to the time-sequence of sensorimotor gains $G^{(t)}$.

Some examples

As an example, consider a simple task first described by Uno, Kawato, and Suzuki (1989) and shown in figure 40.3*A*. The objective is to reach from point T1 to T2. In one condition, the subject is holding a lightweight tool that moves freely in air. In a second condition, the tool is attached to a spring that pulls the hand to the right. Without the spring, people reach in a straight line. This is the path that minimizes the cost. However, once the spring is attached, the straight path incurs

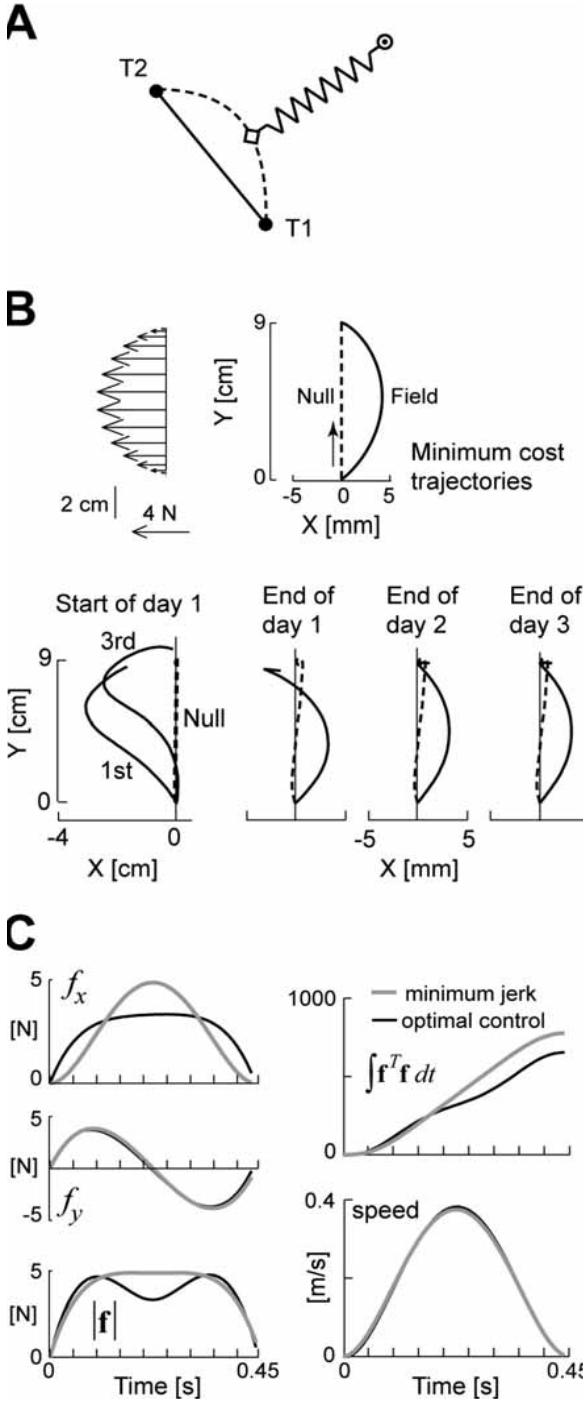


FIGURE 40.3 Task dynamics affect reach trajectories. (*A*) The task is to reach from point T1 to T2. In one condition, the reach takes place in free space (straight line). In another condition, a spring is attached to the hand. In this case, the subject chooses to move the hand along an arc. (*B*) A velocity-dependent force field pushes the hand perpendicular to its direction of motion. For example, for an upward movement, the forces push the hand to the left. The motion that minimizes cost of equation 1 is not a straight line but one that has a curvature to the right. The data show hand paths for a typical subject at start of training on day 1 and then at the end of training each day. Except for the first and third trials, all other trajectories are an average of 50 trials. (*C*) A rationale for why a curved movement is of lower cost. The curves show simulation results on forces that the controller produces and speed of movement in the optimal control scenario of equation 1 and in a scenario where the objective is to minimize jerk. (*A* is redrawn from Uno et al., 1989. Data in parts *B* and *C* are from Izawa et al., 2008.)

substantially more motor costs than a curved path. The curved path is the one that subjects choose (Uno et al., 1989).

In our second example, the task is to move one's hand from one point to another in a given amount of time (450 ms), but now instead of a spring, there is a velocity-dependent field that pushes the hand perpendicular to its direction of motion. Before the field is imposed, the motion that minimizes the cost (and maximizes probability of reward) is simply a straight line with a bell-shaped velocity profile. However, when the field is imposed, the solution is no longer a straight line (Izawa, Rane, Donchin, & Shadmehr, 2008). For example, if the field pushes the hand to the left, the policy that produces the least cost in terms of equation 1 is one that moves the hand slightly to the right of a straight line, resulting in a curved movement that appears to overcompensate for the forces (figure 40.3*B*). As subjects train, their hand paths converge to this curved trajectory. To see the rationale for this behavior, figure 40.3*C* plots the forces produced by the optimal controller and compares it to forces that must be produced if a mass is moving along a "minimum-jerk" trajectory. By moving the hand along a curved path, the optimal controller produces less total force: It overcompensates early into the movement when the field is weak but undercompensates at peak speed when the field is strongest. Therefore, the curved path actually produces less total force than a straight trajectory does. People produce similarly curved trajectories when they move in such fields (Thoroughman & Shadmehr, 2000).

The cerebellum: Predicting sensory consequences of motor commands

According to the theory, we generate motor commands on the basis of beliefs about the state of our body and the environment (equation 5). This state estimate depends on two quantities: a prediction and an observation. The prediction comes from an internal model that uses a copy of the motor commands to estimate the state change that is expected to occur. The observation comes from the sensory system that provides a measure of those state changes. That is, our beliefs are not based on our observations alone. Rather, our beliefs are a combination of what we predicted and what we observed (Kording & Wolpert, 2004a; Vaziri, Diedrichsen, & Shadmehr, 2006).

Some movements are so fast that there is no time for the sensory system to play a role. A prominent example is

control of saccades (rapid eye movements that move the eyes to a new location typically within 50–80 ms). Such movements are too brief for visual feedback to influence saccade trajectory. In fact, the brain actively suppresses visual processing during saccades to reduce the perception of motion (Thiele, Henning, Kubischik, & Hoffmann, 2002). Furthermore, proprioceptive signals from the eyes do not play any significant role in controlling saccade trajectories (Keller & Robinson, 1971; Guthrie, Porter & Sparks, 1983). Thus the brain must guide saccade trajectories in the absence of sensory feedback. How is this accomplished? A plausible solution is for the brain to use an internal estimate of the state of the eye, derived from a copy of ongoing motor commands (Robinson, 1975). This internal feedback probably accounts for the fact that variability at saccade initiation is partially corrected as the saccade progresses (Quaia, Pare, Wurtz, & Optican, 2000). That is, saccades are steered midflight via an internal feedback system (Chen-Harris et al., 2008).

What are the neural substrates of this internal feedback? The available evidence points to the cerebellum (Optican & Quaia, 2002; Optican, 2005). That is, the cerebellum appears to act as a forward model of the plant to produce midflight corrections. A simple experiment can test whether the cerebellum plays a role in predicting consequences of self-generated motor commands. Nowak, Timmann, and Hermsdorfer (2007) asked subjects to hold a force transducer that measures grip force, and then they attached a basket to the transducer. The experimenter dropped a ball into the basket. When the ball dropped, it exerted a downward force on the hand. The subject responded by squeezing the transducer so that it would not slip out of his or her hand. Because there are delays in sensing the impact of the ball, the grip response came about 100 ms after the ball's impact. Nowak and colleagues (2007) described patient HK, who did not have a cerebellum, owing to a very rare developmental condition. When the experimenter dropped the ball into the basket, both the healthy individuals and HK showed the delayed response. Therefore, the sensory feedback pathways appeared to be intact. In a subsequent trial, the subject (rather than the experimenter) dropped the ball. In a healthy individual, the brain can predict that the release of the ball will soon result in an impact that will increase the downward load. In anticipation of this event, the healthy individual squeezed the basket's handle harder around the time when the ball was released. HK, however, could not make this anticipatory adjustment. Rather, she responded to the perturbation in the same way that she responded when the ball was dropped by the experimenter. Therefore, the cerebellum appears to be required for the ability to predict the sensory consequences of motor commands (Wolpert, Miall, & Kawato, 1998).

The cerebellum and construction of internal models

It is not easy to make accurate predictions about the sensory consequences of motor commands; our muscles respond differently depending on their fatigue state, and our limbs move differently depending on whether we are holding a light or heavy object. To maintain accuracy of the predictions, our brain needs to learn from the sensory feedback and adapt its internal model. This adaptation can be simple, such as changing parameter values of a known structure (changing A, B, or H in equation 2), or complex, such as identifying the structure de novo (replacing the linear form of equation 2 with some nonlinear function). Adjusting the parameters of an existing model produces rapid adaptation, whereas learning a new structure is likely to require much practice. The cerebellum appears to be one of the crucial sites for both processes.

Cerebellar damage often prevents individuals from learning how to use novel tools. For example, when subjects are asked to move the handle of a robotic tool to manipulate cursor positions, they may not be able to learn to compensate for forces generated by the robot (Maschke, Gomez, Ebner, & Konczak, 2004; Smith & Shadmehr, 2005) or to compensate for the novel visual feedback through a mirror (Sanes, Dimitrov, & Hallett, 1990). If the cerebellum is the crucial site for learning internal models, then it probably makes its contribution to control of reaching via its outputs to the thalamus, which in turn projects to the cerebral cortex. In humans, it is possible to reversibly disrupt this pathway. Essential tremor patients are occasionally treated with deep-brain stimulators that artificially disrupt the ventrolateral thalamus, improving their tremor. However, these patients learn the reach task better when the stimulator is turned off (Chen, Hua, Smith, Lenz, & Shadmehr, 2006). In contrast, patients with damage to the basal ganglia showed little or no deficit in adaptation with either the robot task (Smith & Shadmehr, 2005) or the mirror task (Agostino, Sanes, & Hallett, 1996; Gabrieli, Stebbins, Singh, Willingham, & Goetz, 1997). Therefore it seems quite likely that the cerebellum is a key structure that allows us to learn tool use.

Experiments show that the cerebellar damage causes abnormalities in adaptation to both kinematic (Tseng, Diedrichsen, Krakauer, Shadmehr, & Bastian, 2007) and force (Smith & Shadmehr, 2005) perturbations. One unifying concept is that the cerebellum may be the site of the internal model that predicts the sensory consequences of motor commands (equation 2). The output of the internal model could be used to generate a prediction error that drives adaptation and also be used to update a previous estimate of limb state. Support for this idea comes from a recent experiment in which transcranial magnetic stimulation was used to disrupt the lateral cerebellum in human subjects while they slowly

moved their arm in preparation for making a rapid reaching movement (Miall, Christensen, Owen, & Stanley, 2007). Reaching errors in initial direction and final finger position suggested that the reaching movements had been made from an estimated hand position that was approximately 140 ms out of date, consistent with a role for the cerebellum in iteratively updating limb state.

Learning the rewarding nature of sensory states

You might expect that a severely amnesic individual who was performing a novel task would have to be regularly reminded of the task's instructions. For example, if it is a reaching task, we might have to repeat "try to move the cursor to the target fast enough so it explodes." However, when we examined the severely amnesic patient HM on the standard reach adaptation task with the robot (Shadmehr, Brandt, & Corkin, 1998), after he had exploded a few targets, he no longer needed verbal reminders. The visual appearance of the target was enough for him to initiate a reaching movement. Strikingly, when he returned a few hours later (or the next day), he voluntarily reached for the robot handle and began preparing for onset of targets by moving the cursor to the center location (naïve individuals avoid touching the machine). It was clear that despite having no conscious recollection of having done the task before, some part of HM's brain recognized that the contraption was a tool that had a particular purpose: to manipulate cursors on a screen. This behavior suggested that during the first session, he implicitly learned the reward basis of the task (equation 1). (For HM, the target explosion triggered a childhood memory of going bird hunting. As he was performing the task and was able to get a target explosion, he would spend the next few minutes describing the memory in detail: the type of gun that he used, the porch in the rear of his childhood home, the terrain of the woods in his backyard, and the kinds of birds that he hunted.) What brain regions were involved in learning the rewarding nature of bringing the cursor to the target?

Experiments on action selection in rodents provide important insights into this question. For example, suppose that a rat is released into a pool of water from some random starting point. A platform is positioned in a specific location just below the water line and cannot be seen. The platform is always at the same location in the pool. Rats dislike being wet and will try to find a way to elevate themselves. The normal rat can learn to locate the platform position by paying attention to the visual cues that surround the pool. This requires learning a spatial map of where the platform is located with respect to the surrounding visual cues. With repeated swims, the animal learns a spatial map. This spatial map is analogous to a reward function that associates places in the pool with the likelihood of the platform (and therefore the likelihood of not having to be wet).

Once the map has been learned, the animal can find the platform regardless of where the rat is released into the water because the map is with respect to the cues on the walls. If the platform is removed, the normal animal will spend most of the time searching in the region where the platform should be. Sometimes, certain cues are rewarding no matter where they are located. Consider a pool where there are two hidden platforms: one that is large enough for the rat to mount and one that is too small. Both have a distinct visual cue associated with them: a little flag attached to each platform, each of a different color, sticking out of the water. Suppose that the flag attached to the large platform is red and the flag attached to the small platform is green. The platforms may be positioned in any part of the pool and will change from trial to trial. Therefore, in this experiment, the animal needs to learn that the red flag indicates the location of the suitable platform and is a rewarding object. In another version of the experiment, the large platform will always be located in a particular spatial location, but the flag on top of it will be a random color. In this version of the experiment, the animal needs to learn that it is not the color of the flag that is important, but the spatial location.

We see that there is a natural competition between the learning systems that might be involved in these two conditions: Is the platform in the same "place" as before (where place refers to a location in the spatial map), or is the platform always where the red flag is located? Packard and McGaugh (1992) performed both experiments by having their animals swim eight times per day for a number of days. They recorded the number of times the animals mounted the small platform and labeled these as errors. In the first experiment, in which reward was associated with the red flag, healthy animals gradually learned to swim to the red flag. Interestingly, animals with damage to the medial temporal lobe learned the task just as well as the healthy controls did. However, animals with damage to the caudate nucleus were much slower in learning the association. After days of training, they continued to attempt to mount the platform under the green flag. Therefore it appears that the ability to associate reward to stimuli regardless of its spatial location depends on the basal ganglia.

In the second experiment, in which reward was associated with a spatial location, healthy animals gradually learned to swim to that location and ignore the color of the flag. Animals with damage to the caudate nucleus performed similarly to the healthy controls. However, animals with damage to the medial temporal lobe were much slower in learning the association. Therefore the ability to associate reward to a spatial location depends on the medial temporal lobe.

Returning to our observations in HM, we would speculate that it was his basal ganglia that learned that if he were to place the cursor in the box on the screen and do so rapidly, a rewarding state would be experienced (explosions, which triggered a pleasant childhood memory). During the later sessions, the visual appearance of the machine and the act of holding its handle likely triggered a recall of this reward structure.

Effects of striatal damage on the assessment of movement costs and rewards

One of the striking features of damage to the human striatum is micrographia, an impairment of writing in which letters become very small and writing speed becomes slow. This condition is most common in degenerative diseases of the basal ganglia such as Parkinson's disease (Van Gemmert, Teulings, & Stelmach, 2001). However, it can also occur with focal lesions. Consider patient FF, an individual who suffered an ischemic stroke in the left basal ganglia, in the head of the caudate nucleus and the anterior part of the putamen (Barbarulo, Grossi, Merola, Conson, & Trojano, 2007). When FF was asked to copy a four- or eight-letter string of characters, writing with the right hand was much smaller than with the left hand. Micrographia reflects an abnormal choice of speed and amplitude and is one manifestation of generalized slowing of movement (bradykinesia).

In the optimal control framework, there are no desired trajectories for our movements. Rather, the path is a result of a control policy (equation 5), which itself is a result of minimization of a cost (equation 1). The cost depends on two quantities: spatial accuracy (error cost) and required effort (energy cost). Accuracy requirements influence speed selection, due to the signal-dependent noise property of motor commands. The desired accuracy of a movement sets an upper limit on the maximum speed of a movement. The accuracy term of the cost function offers an explanation for the wealth of experimental data demonstrating speed-accuracy tradeoff in reaching movements. Normal movements, however, do not appear to be made at the limits imposed by the speed-accuracy tradeoff: We can reach for an object faster than usual without appreciable loss of accuracy. Although very little experimental data exist on spontaneous speed selection, the effort term of the cost function offers a potential explanation for this phenomenon; that is, perhaps micrographia is an indication of an abnormally high motor cost.

One of us recently tested this idea that in Parkinson's disease, there may be an abnormally high cost associated with motor commands (Mazzoni, Hristova, & Krakauer, 2007). We required healthy control subjects to make accurate reaching movements of specified speeds. As the required speed increased, subjects took longer (required more trials) to accumulate a set number of movements at the required speed. This reluctance to move faster could be explained by the increase in required energy as well as by the degradation of spatial accuracy and thus did not disambiguate the contribution of these two costs. We then compared the performance of patients with Parkinson's disease to that of control subjects in this task. Parkinson's disease patients demonstrated normal spatial accuracy in each condition but required more trials than controls to accumulate the required number of movements in each speed range. The patients' increased reluctance to execute movements requiring greater effort, in spite of preserved spatial accuracy, provided experimental demonstration of the contribution of energy cost to speed selection, independent of spatial accuracy. The implication is that bradykinesia results when striatal dysfunction changes the value of effort minimization (increased sensitivity to effort cost; L in equation 1) relative to that of accuracy optimization (error cost; Q in equation 1). Thus it appears that the basal ganglia either provides the motor motivation signal, which is then used to compute the cost-to-go elsewhere, or is where the cost-to-go is computed.

Parietal cortex damage and state estimation

Sometimes goal states change as the task is being performed. For example, when one reaches to pick up a pen, the pen may start rolling away. Healthy individuals have no problems adjusting their movements to compensate for this change. However, parietal patients show particular difficulties with this task. For example, if parietal damage impairs representation of visual states contralateral to the fixation, then motion of the goal state to this region during a movement impairs the ability to adjust the reach mid-flight. Grea and colleagues (2002) observed this phenomenon in a patient with bilateral posterior parietal cortex damage. The patient had no problems reaching to targets in central fixation. However, when the target shifted to the right at reach onset, the subject continued to reach to the original location of the target as if the target had not moved.

Disruption of the parietal cortex in healthy individuals can produce a similar phenomenon. Desmurget and colleagues (1999) provided a single pulse via a transcranial magnetic stimulator as the reach to the target began. On trials in which the target jumped, most of the participants had hand movements that disregarded the shift in the target location.

Let us examine these results in the framework of figure 40.2. The relevant state variables in this task include position of the limb (in proprioceptive and visual coordinates) and the position of the target (in visual coordinates). As motor commands are generated, the forward model should update its predicted state of the limb. Generally, we expect targets

to remain stationary, and therefore the output of the forward model should continue to predict the target position. Together, these predictions represent the prior belief about the state of the body and the world. The sensory feedback from proprioception and vision is integrated with this prediction to make a posterior belief. When the target jumps, the novel sensory information needs to be integrated with the output of the forward model. If it is not, the reach will continue to the prior expectation of its location. The results noted above suggest that either this integration step is affected by damage or stimulation of the parietal cortex or that the sensory information outside the central fixation region cannot reach the integration step.

Limitations in applying the theory to biological motor control

This review of motor control has been written within the framework of optimal feedback control. At the heart of the theory is the conjecture that animals make voluntary movements in order to acquire the most reward while expending the least effort. However, the theory cannot make a behavioral prediction unless we can specify three kinds of information: (1) what the costs and rewards are; (2) what the constraints are, that is, dynamics of the task; and (3) what the mechanisms of state estimation are. In this review, we have chosen a specific set of equations to represent each kind of information. However, it is not difficult to find examples of behavior that are inconsistent with our formulation.

The cost that we wrote in equation 1 is perhaps the simplest possible cost function for goal-directed movements. How seriously can we take this specific representation? As demonstrated by attempts to reverse-engineer the cost (Kording & Wolpert, 2004b), the quadratic cost function should not be taken too seriously.

Consider a set of experiments that highlighted the importance of costs associated with postural stability, a quantity that we did not include in equation 1. Scheidt and Ghez (2007) explored a task in which continuous random noise perturbed the hand at rest. This constraint encouraged increasing the cocontraction levels of muscles. However, the noise was present only during the postural phase of the task and disappeared when subjects made a reaching movement. They found that if a kinematic perturbation required adaptation of the movement, the learning did not generalize to the postural phase at the end of the movement. They suggested that the control processes that moved the limb appeared distinct from control processes that set muscle activity levels during posture. If so, do these processes have separate costs? A recent study suggests that the answer is yes, the weighting of postural cost is flexible and can be determined by task context (Liu & Todorov, 2007).

Finally, consider an experiment by Jax and Rosenbaum (2007) in which they asked subjects to make arm movements to an array of 12 targets positioned in a 16-cm radius circle on a vertical screen. Targets were presented randomly, and in some trials, an obstacle was presented halfway between the start and the target. The same target was never shown twice in a row. Interestingly, whenever a no-obstacle trial followed an obstacle trial, subjects made curved rather than straight trajectories. However, the movements straightened out when a no-obstacle trial followed another no-obstacle trial. Why make a suboptimal curved trajectory when you see that there is no obstacle?

These results highlight a number of important problems with our framework. First, without knowing precisely the costs and rewards of a movement, it will not be possible to make quantitatively reliable predictions of behavior. Without a priori predictions, how can the theory be falsified?

Second, what are the timescales of optimization? Is optimization computed in the reaction time of each trial de novo? The timescale appears to be longer than a single trial, as exemplified by the example from Jax and Rosenbaum (2007). Certainly, new costs can be conjured up. For example, in this case, we can assume that finding feedback control gains that minimize a cost requires neural processing that itself has a cost, so it might be more efficient to allow the solution in one trial to linger on to influence the solution in the next trial. Or perhaps there is a cost in switching control policies.

Third, what is the timescale of system identification? Our body changes over multiple timescales. Muscles fatigue and recover quickly, objects are lifted and replaced rapidly, yet aging can produce gradual loss of motor neurons and transformation of muscle fibers. In other words, the parameters of the constraint equation and perhaps its structure are changing over multiple timescales. Unfortunately, we cannot make optimized movements unless we have an accurate set of constraint equations, that is, an accurate internal model. When we see a suboptimum movement, can we dissociate the effects of an inaccurate internal model from effects of an inaccurate cost function?

Finally, what is the alternative hypothesis to this theory? At this time, the alternative is another cost or constraint, not a fundamentally distinct theory. However, formalization of a theory is the key step that accelerates its evolution toward acceptance or rejection.

Conclusions

The relationship between theories and the neural machinery that implements them is still in the courtship stage, but despite the separation, it has begun to bear modest fruit; theories have informed the neural basis of motor control in patients, while lesion studies have informed the algorithms

and representations that implement the computational theories. The result is the functional anatomy of voluntary movements outlined in figure 40.2*B*. In this framework, a role for the cerebellum is system identification, that is, predicting the changes in state that arise as a result of motor commands. A role for the parietal cortex is state estimation, in which predictions about sensory feedback are integrated with visual and proprioceptive observations to form beliefs about states of our selves and objects/people around us. The basal ganglia may play a role in computing a cost-to-go function, estimating value of states and costs of motor commands. Finally, once a goal state has been selected, motor cortical areas minimize this cost function and transform state estimates into motor output by formulating a feedback control policy.

ACKNOWLEDGMENTS The work was supported by National Institutes of Health (NIH) grants K02-048099 and R01-052804 to JWK and R01-037422 to RS.

REFERENCES

Agostino, R., Sanes, J. N., & Hallett, M. (1996). Motor skill learning in Parkinson's disease. *J. Neurol. Sci.*, *139*, 218–226.

Barbarulo, A. M., Grossi, D., Merola, S., Conson, M., & Trojano, L. (2007). On the genesis of unilateral micrographia of the progressive type. *Neuropsychologia*, *45*, 1685–1696.

Chen, H., Hua, S. E., Smith, M. A., Lenz, F. A., & Shadmehr, R. (2006). Effects of human cerebellar thalamus disruption on adaptive control of reaching. *Cereb. Cortex*, *16*, 1462–1473.

Chen-Harris, H., Joiner, W. M., Ethier, V., Zee, D. S., & Shadmehr, R. (2008). Adaptive control of saccades via internal feedback. *J. Neurosci.*, *28*, 2804–2813.

Desmurget, M., Epstein, C. M., Turner, R. S., Prablanc, C., Alexander, G. E., & Grafton, S. T. (1999). Role of the posterior parietal cortex in updating reaching movements to a visual target. *Nat. Neurosci.*, *2*, 563–567.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.*, *47*, 381–391.

Gabrieli, J. D. E., Stebbins, G. T., Singh, J., Willingham, D. B., & Goetz, C. G. (1997). Intact mirror-tracing and impaired rotary-pursuit skill learning in patients with Huntington's disease: Evidence for dissociable memory systems in skill learning. *Neuropsychology*, *11*, 272–281.

Grea, H., Pisella, L., Rossetti, Y., Desmurget, M., Tilikete, C., Grafton, S., Prablanc, C., & Vighetto, A. (2002). A lesion of the posterior parietal cortex disrupts on-line adjustments during aiming movements. *Neuropsychologia*, *40*, 2471–2480.

Guthrie, B. L., Porter, J. D., & Sparks, D. L. (1983). Corollary discharge provides accurate eye position information to the oculomotor system. *Science*, *221*, 1193–1195.

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*, 780–784.

Izawa, J., Rane, T., Donchin, O., & Shadmehr, R. (2008). Motor adaptation as a process of reoptimization. *J. Neurosci.*, *28*, 2883–2891.

Jax, S. A., & Rosenbaum, D. A. (2007). Hand path priming in manual obstacle avoidance: Evidence that the dorsal stream does not only control visually guided actions in real time. *J. Exp. Psychol. Hum. Percept. Perform.*, *33*, 425–441.

Jones, K. E., Hamilton, A. F., & Wolpert, D. M. (2002). Sources of signal-dependent noise during isometric force production. *J. Neurophysiol.*, *88*, 1533–1544.

Keller, E. L., & Robinson, D. A. (1971). Absence of a stretch reflex in extraocular muscles of the monkey. *J. Neurophysiol.*, *34*, 908–919.

Kording, K. P., & Tenenbaum, J. B., Shadmehr, R. (2007). The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nat. Neurosci.*, *10*, 779–786.

Kording, K. P., & Wolpert D. M. (2004a). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–247.

Kording, K. P., & Wolpert D. M. (2004b). The loss function of sensorimotor learning. *Proc. Natl. Acad. Sci. USA*, *101*, 9839–9842.

Liu, D., & Todorov, E. (2007). Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *J. Neurosci.*, *27*, 9354–9368.

Maschke, M., Gomez, C. M., Ebner, T. J., & Konczak, J. (2004). Hereditary cerebellar ataxia progressively impairs force adaptation during goal-directed arm movements. *J. Neurophysiol.*, *91*, 230–238.

Mazzoni, P., Hristova, A., & Krakauer, J. W. (2007). Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *J. Neurosci.*, *27*, 7105–7116.

Miall, R. C., Christensen, L. O. D., Owen, C., & Stanley, J. (2007). Disruption of state estimation in the human lateral cerebellum. *PLoS Biol.*, *5*, e316.

Nowak, D. A., Timmann, D., & Hermsdorfer, J. (2007). Dexterity in cerebellar agenesis. *Neuropsychologia*, *45*, 696–703.

Optican, L. M. (2005). Sensorimotor transformation for visually guided saccades. *Ann. NY Acad. Sci.*, *1039*, 132–148.

Optican, L. M., & Quaia, C. (2002). Distributed model of collicular and cerebellar function during saccades. *Ann. NY Acad. Sci.*, *956*, 164–177.

Packard, M. G., & McGaugh, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behav. Neurosci.*, *106*, 439–446.

Quaia, C., Pare, M., Wurtz, R. H., & Optican, L. M. (2000). Extent of compensation for variations in monkey saccadic eye movements. *Exp. Brain Res.*, *132*, 39–51.

Robinson, D. A. (1975). Oculomotor control signals. In P. Bachy-Rita & G. Lennerstrand (Eds.), *Basic mechanisms of ocular motility and their clinical implications* (pp. 337–374). Oxford, UK: Pergamon.

Sanes, J. N., Dimitrov, B., & Hallett, M. (1990). Motor learning in patients with cerebellar dysfunction. *Brain*, *113*, 103–120.

Scheidt, R. A., & Ghez, C. (2007). Separate adaptive mechanisms for controlling trajectory and final position in reaching. *J. Neurophysiol.*, *98*, 3600–3613.

Shadmehr, R., Brandt, J., & Corkin, S. (1998). Time dependent motor memory processes in H.M. and other amnesic subjects. *J. Neurophysiol.*, *80*, 1590–1597.

Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.*, *14*, 3208–3224.

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.*, *4*, e179.

Smith, M. A., & Shadmehr, R. (2005). Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *J. Neurophysiol.*, *93*, 2809–2821.

Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H., & Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. *Exp. Brain Res.*, *142*, 284–291.

Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K. P. (2002). Neural mechanisms of saccadic suppression. *Science*, *295*, 2460–2462.

Thoroughman, K. A., & Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature*, *407*, 742–747.

Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.*, *5*, 1226–1235.

Trommershauser, J., Gepshtein, S., Maloney, L. T., Landy, M. S., & Banks, M. S. (2005). Optimal compensation for changes in task-relevant movement variability. *J. Neurosci.*, *25*, 7169–7178.

Tseng, Y. W., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *J. Neurophysiol.*, *98*, 54–62.

Uno, Y., Kawato, M., & Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement: Minimum torque-change model. *Biol. Cybern.*, *61*, 89–101.

Van Gemmert, A. W., Teulings, H. L., & Stelmach, G. E. (2001). Parkinsonian patients reduce their stroke size with increased processing demands. *Brain Cogn.*, *47*, 504–512.

Vaziri, S., Diedrichsen, J., & Shadmehr, R. (2006). Why does the brain predict sensory consequences of oculomotor commands? Optimal integration of the predicted and the actual sensory feedback. *J. Neurosci.*, *26*, 4188–4197.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci.*, *2*, 338–347.

41 Forward Models and State Estimation in Posterior Parietal Cortex

GRANT H. MULLIKEN AND RICHARD A. ANDERSEN

ABSTRACT During on-line control of movement, the posterior parietal cortex (PPC) serves as a functional bridge between sensory and motor areas in the brain. One of the sensorimotor functions of this area appears to be prediction of the state of the arm during movement. Because sensory information is substantially delayed, it has been proposed that the brain makes use of an internal forward model that integrates both sensory and motor feedback signals to estimate current and upcoming positions and motions of the limb during reaching. These predicted states are more useful for rapid on-line control than are delayed sensory signals. The first part of this chapter focuses on investigations of on-line control mechanisms in PPC. The results of these studies indicate that one of the functions of PPC is to serve as a forward model. The second section highlights research that aims to read-out forward state estimates from PPC neurons and harness them for direct control of neural prostheses.

A growing body of clinical and psychophysical evidence supports the theory that the brain makes use of an internal model during control of movement; a sensorimotor representation of the interaction of one's self with the physical world (Jordan, 1995; Kawato, Furukawa, & Suzuki, 1987). Two primary types of internal models for sensorimotor control have been proposed: the forward model and the inverse model. A forward model (i.e., forward output model) predicts the sensory consequences of a movement (Jordan & Rumelhart, 1992; Miall & Wolpert, 1996; Wolpert, Ghahramani, & Jordan, 1995). That is, it mimics the behavior of a motor system by predicting the expected, upcoming state of an end effector (e.g., sensory feedback of one's own limb) using knowledge of the characteristic dynamics of the system as well as stored copies of recently issued motor commands. Conversely, an inverse model encodes the motor commands necessary to produce a desired outcome (Atkeson, 1989). That is, an inverse model estimates the set of procedures (e.g., motor commands) that will cause a particular state of the motor system to occur. While inverse models likely play an important role in sensorimotor control, they will not be discussed further in this chapter; instead, we will place emphasis on the forward model and, in particular, the role of the posterior parietal cortex (PPC) in forward state estimation for motor planning and control.

GRANT H. MULLIKEN Computational and Neural Systems, California Institute of Technology, Pasadena, California
RICHARD A. ANDERSEN Computation and Neural Systems, Division of Biology, California Institute of Technology, Pasadena, California

Movement intention and anticipation in PPC

PPC is a critical node for bridging sensory and motor representations in the brain. PPC associates multiple sensory modalities (e.g., visual—the dominant sensory input to PPC, somatosensory, and auditory) and transforms these inputs into a representation that is useful for guiding actions to objects in the external world (Andersen & Buneo, 2002). Evidence from lesions studies indicates that damage to PPC results in an inability to link the sensory requirements of a task with the appropriate motor behavior necessary to complete it. For example, parietal lesion patients can have difficultly planning skilled movements, a condition known as apraxia (Geshwind & Damasio, 1985). Impairments from apraxia can range from an inability to properly perform an instructed or desired arm movement to how to coordinate a specific sequence of movements to accomplish an end goal.

Numerous neurophysiological studies in monkeys have shed light on the neural correlates of reach planning in PPC. Monkeys have served as a successful model for studying sensorimotor representations in humans since the two species engage in a variety of similar sensorimotor behaviors. Moreover, functional magnetic resonance imaging (fMRI) studies have provided evidence that PPC's functional role is similar in both monkeys and humans (Connolly, Andersen, & Goodale, 2003; DeSouza et al., 2000; Pellijeff, Bonilha, Morgan, McKenzie, & Jackson, 2006; Rushworth, Paus, & Sipila, 2001). When trained monkeys plan a reach to an illuminated target, the firing rates of neurons in the medial bank of the intraparietal sulcus (MIP) generally reflect a combination of both sensory and motor parameters

(Andersen & Buneo, 2002; Mountcastle, Lynch, Georgopoulos, Sakata, & Acuna, 1975; Robinson, Goldberg, & Stanton, 1978). Importantly, during a memory period in which the monkey must maintain a reach plan to the remembered location of an extinguished target, elevated neural activity persists in PPC before the reach is executed, suggesting that these neurons likely encode the *intention* to reach rather than the visual stimulus location (Snyder, Batista, & Andersen, 1997). Furthermore, neural responses in MIP are generally correlated more strongly with the motor goal, and not the visual cue, during antireach paradigms in which the target cue direction is dissociated from the reach direction (Eskandar & Assad, 1999; Gail & Andersen, 2006; Kalaska & Crammond, 1995).

PPC is a reasonable location for a forward model of the arm to reside (which would predict the sensory consequences of an upcoming arm movement command) given its substantial reciprocal connections with downstream motor areas (Johnson, Ferraina, Bianchi, & Caminiti, 1996; Jones & Powell, 1970). Along these lines, many researchers have suggested that the "early" discharge of neurons in area 5 prior to initiation of an arm movement might reflect the integration of an efference copy signal fed back to PPC from frontal motor areas (Kalaska, Caminiti, & Georgopoulos, 1983; Seal, Gross, & Bioulac, 1982). Interestingly, Seal and colleagues also showed that area 5 responses that occurred prior to movement onset were generally not sensory in origin and further demonstrated that these early responses persisted even after deafferentation. However, some caution should be advised in attempting to infer the causal flow of information in parietofrontal circuits during reach preparation using single-area correlation analyses. For instance, it is possible that planning and forward model prediction (which relies on efference copy) may be carried out by distinct neural processes within PPC. Future simultaneous multiarea recordings, combined with microstimulation approaches, may help to shed light on the directional flow of information in these recurrent interarea circuits during movement preparation.

A Forward Model for Eye Position PPC is also a possible candidate for a forward model of eye position, since a variety of eye behavior–related signals, such as saccade and fixation responses, have been described in this region (Mountcastle et al., 1975). Area 7a saccade responses begin largely after a saccade occurs, while lateral intraparietal (LIP) saccade responses can occur before, during, or after saccades (Andersen, Essick, & Siegel, 1987). Interestingly, Duhamel, Colby, and Goldberg (1992) showed that the receptive fields (RFs) of neurons in LIP can update their receptive fields before an eye movement occurs. Forty-four percent (16 out of 36) of their LIP sample anticipated the sensory outcome of an impending saccade (i.e., a stimulus appearing in the future location of the RF), and adjusted their responses approximately 80 ms before the saccade was launched. It is conceivable that this predictive updating relies upon a forward model of eye position within PPC, which estimates the upcoming eye position from oculomotor commands, though direct evidence of the anticipatory eye position signal itself in PPC has not been reported. An eye position signal in PPC could potentially be derived from passive sensory feedback from the eye muscles (Wang, Zhang, Cohen, & Goldberg, 2007) and/or the integration of saccade command signals. It would be interesting to see whether a component of the eye position signal in PPC might also encode anticipatory information (ahead of passive sensory feedback) about the current state of the eye position during fixations between saccades. Evidence of updating RFs has also been reported in other brain areas; therefore it is quite possible that multiple regions are involved in encoding a forward model of the state of the eye. For instance, Sommer and Wurtz (2006) discovered a feedback circuit in which the superior colliculus conveys a copy of the oculomotor command (i.e., corollary discharge) to the frontal eye field (FEF), which they showed to be necessary for accurate updating of RFs in the FEF. Last, response field updating neurons in PRR, which predominantly encode an intended reach direction in eye-centered coordinates, update their response fields when an intervening saccade occurs and thereby maintaining an eye-centered motor plan even when gaze is shifted (Batista, Buneo, Snyder, & Anderson, 1999; Buneo, Jarvis, Batista, & Andersen, 2002). It would be interesting to test whether the reach response fields of these PRR neurons also exhibit anticipatory updating just before the eye moves, similar to the cells found in LIP by Duhamel and colleagues (1992).

Reafference Cancellation in PPC A forward model's ability to predict the sensory consequences of an action is useful to an organism because a given sensory outcome can be produced by a variety of potential causes (Claxton, 1975; Cullen, 2004; Poulet & Hedwig, 2003; Roy & Cullen, 2004; Sperry, 1950; Weiskrantz, Elliott, & Darlington, 1971). In particular, the output of a forward model can be used as an internal reference signal to cancel the sensory effects of self-motion. For example, motion on our retina can occur because of movement in the physical world (afference) or because of motion induced by an eye movement itself (reafference). Therefore to correctly perceive the motion of an external stimulus, the brain must distinguish afferent motion from reafferent motion. A subtractive comparison between a forward model's estimate of the expected sensory outcome of an eye movement and the actual sensory signals could remove this retinal shift from our perception. For example, such an internal reference signal is used for perceptual stability during smooth-pursuit eye movements

(Bradley, Maxwell, Andersen, Banks, & Shenoy, 1996; Haarmeier, Bunjes, Lindner, Berret, & Thier, 2001). Interestingly, clinical evidence presented by Haarmeier, Thier, Repnow, and Peterson (1997) suggested that parieto-occipital regions may be involved in performing the comparison between self-induced and external sensory motion during smooth-pursuit eye movements.

Reafference generation and comparison mechanisms are also likely employed for the perception of arm movements, for example, to distinguish self-generated arm movement from movement in the environment (e.g., the movement of others) and/or the movement of one's arm by an external force. Positron emission tomography (PET) imaging studies have provided evidence supporting PPC's role in reconciling intentions with sensory consequences. For instance, Fink and colleagues (1999) displayed nonveridical visual feedback of a subject's left hand by displaying a mirror image of a subject's right hand while they performed a bimanual coordination task. Such incongruent visual feedback resulted in an increase in bilateral PPC activation (area 40 and area 7) as well as bilateral dorsal prefrontal cortex activation. Later, Farrer and colleagues (2003) performed an experiment in which they systematically manipulated the degree of control with which subjects were able to perform a joystick task by perturbing visual feedback of their hand movements, rotating the direction of the virtual hand movement by a variable amount (i.e., 25°, 50°, and a condition with no correspondence). They found a graded activation in the inferior parietal lobule, such that regional cerebral brain flow increased with decreasing levels of control felt by the subject.

Lesion studies have shown that damage to PPC can lead to deficits in the attribution of agency. For example, Sirigu, Daprati, Pradat-Diehl, Franck, and Jeannerod (1999) showed that apraxia patients with left parietal lesions have a greater tendency to confuse their own movements with the movements of an experimenter. When patients' visual feedback was substituted with the hand movements of an experimenter (who attempted to perform the same movement), patients were more likely to confuse their right hand with the "alien" hand (19% correct ownership judgment), relative to normal control subjects (79% correct). Since the patient's intention and the outcome of the experimenter's movement were largely congruent, patients needed to detect subtle spatiotemporal discrepancies between the time-varying state of the expected state of their hand and the virtual hand on the computer screen. The authors suggested that these deficits were caused by damage to an internal model, which maintains a time-varying representation of a movement in space. In a related study, MacDonald and colleagues tested whether transient disruption of PPC using transcranial magnetic stimulation (TMS) could affect subjects' ability to determine the agency of an observed movement (MacDonald & Paus, 2003). Specifically, they introduced a lag time into the display of the visual feedback of the subject's hand on the computer screen and asked subjects to detect trials in which they perceived a delay between the onset of their own hand movement and the onset of the virtual movement on the screen. Interestingly, the researchers found that during self-generated movement, TMS impaired subjects' ability to detect asynchrony between the onset of actual and virtual hand movements. In contrast, when subjects' hands were passively moved without prior notice to the subject, their judgments were not significantly impaired from pre-TMS control conditions. These results suggest that PPC maintains a time-dependent representation of action that relies upon anticipatory mechanisms (and not only sensory feedback), such as a forward model, to update the state of the arm. This internal state representation is important for making decisions about both the temporal state and the attribution of agency of a movement.

Forward state estimation for on-line control

During execution of a goal-directed arm movement to continuously guide the arm to a target, the brain must maintain an estimate of the time-varying state of the arm (e.g., position and velocity of the arm, coded in a variety of potential coordinate frames) and compare that state measurement with the desired state of the movement. Unfortunately, the human brain, in particular PPC, does not have direct access to the true state of the arm owing to delayed and noise-corrupted measurements of the state from the visual and proprioceptive domains; for example, visual signals typically reach sensorimotor association areas of cortex after a delay of approximately 90 ms (Raiguel, Xiao, Marcar, & Orban, 1999), or 30 ms in the case of proprioception (Petersen, Christensen, Morita, Sinkjaer, & Nielsen, 1998). Subsequent processing delays are incurred during control, owing to sensorimotor integration, motor command generation, and execution, resulting in an average loop delay of more than 100 ms for proprioceptive control (Flanders & Cordo, 1989) and over 200 ms for visuomotor control (Georgopoulos, Kalaska, & Massey, 1981; Miall, Weir, Wolpert, & Stein, 1993). These long delay times severely limit a feedback control system's ability to make rapid adjustments to an ongoing movement and thus increase the likelihood that a reach trajectory might become erroneous and/or unstable.

The Observer Framework Fortunately, the brain can also monitor recently issued motor commands (i.e., efference copy), which can be transmitted centrally (e.g., from frontal motor areas) with little delay time (e.g., one synapse + transmission time < 10 ms) and used by a forward model to form an estimate of the current or upcoming state of the arm well in advance of late-arriving sensory information.

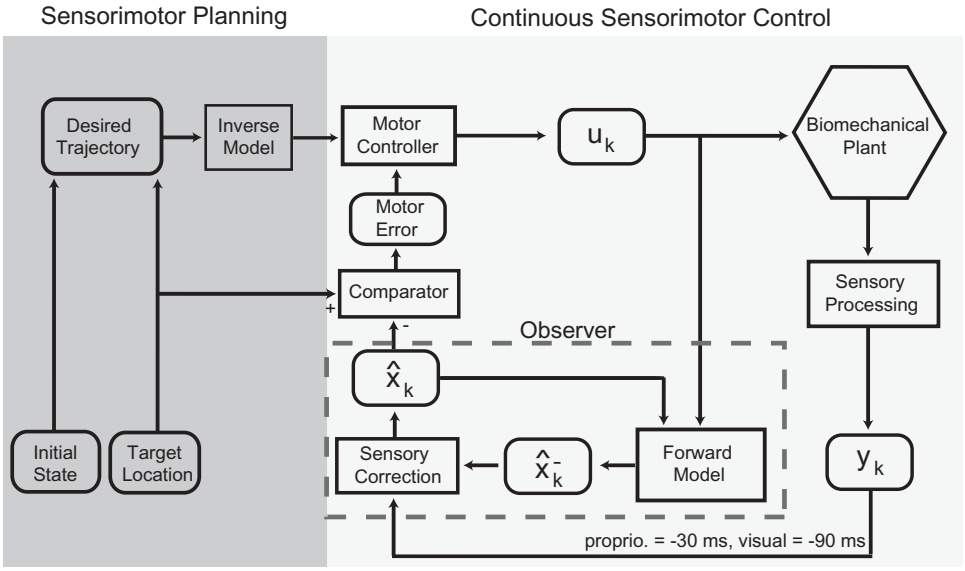


FIGURE 41.1 Flow diagram illustrating sensorimotor integration for reach planning and on-line control. Items in rounded boxes denote pertinent sensorimotor variables; computational processes are contained in rectangular boxes. Prior to a reach, an intended trajectory is formulated as a function of both the initial state of the arm and the desired endpoint, the target location. An inverse model is used to determine a set of motor plans that will result in the desired trajectory. Motor plans are then issued (e.g., by primary motor cortex, M1) and subsequently executed by muscles acting upon the physical environment (i.e., biomechanical plant hexagon). Following movement onset, the state of the arm is continuously monitored and corrected, if necessary, to ensure successful completion of the reach. Critical to rapid on-line correction of movement is the forward model, which generates an anticipatory, a priori estimate of the next state of the arm, $\hat{x}_k$, as a function of the previous state and efference copy. Intermittent sensory feedback is used to refine the a priori estimate of the forward dynamics model (observer). The a posteriori current state estimate, $\hat{x}_k$, can then be evaluated to make corrections to subsequent motor commands. (After Desmurget & Grafton, 2000.)

Since the output of the forward model reflects a best guess of the next state of the arm, errors due to various sources of noise will inevitably accumulate over time for this estimate. Therefore it is likely that sensory observations, which arrive at later times, are also continually integrated by the brain to update and refine the estimate of the forward model (Miall & Wolpert, 1996) (figure 41.1). A system that estimates the state of a movement by combining the output of a forward model with sensory feedback about the state is generally referred to as an observer (Goodwin & Sin, 1984). For linear systems in which the noise is additive and Gaussian, the optimal (i.e., in the mean squared error sense) observer is known as a Kalman filter (Kalman, 1960). Wolpert and colleagues first applied the Kalman filter to model how subjects estimate the sensorimotor state of the hand during goal-directed reaches. They showed that a Kalman filter could accurately account for subjects' estimates of the perceived end location of their hand while making arm movements in the dark (Wolpert et al., 1995). Therefore the Kalman filter can serve as a useful theoretical model for studying sensorimotor state estimation in the brain.

Two linear stochastic equations govern the basic operation of the Kalman filter:

$$x_k = A_k x_{k-1} + B u_{k-1} + w_{k-1} \quad \text{(forward model)} \tag{1}$$

$$y_k = H_k x_k + v_k \quad \text{(state observation model)} \tag{2}$$

where x_k is the time-varying state of the arm at time step k and is modeled as a linear function of the previous state, x_{k-1}, and the control term, u_{k-1}. The control term is considered to be a known motor command, which is likely specified by frontal motor areas (e.g., primary motor or premotor cortex) and then fed back to sensorimotor circuits performing state estimation. For instance, the motor command at each time step might be determined by using an optimization procedure that minimizes a cost function associated with carrying out a particular trajectory (Todorov, 2006). Here, y_k is a sensory measurement (visual and proprioceptive) made at time step k. (Note that sensory feedback is in fact a delayed representation of the state of the arm.)

To estimate the state of the arm at each time step k, the output of the forward model, $\hat{x}_k^-$ (i.e., the a priori estimate), is linearly combined with the difference between the output of the observation model (i.e., the predicted sensory measurement) and the actual sensory measurement. This discrepancy, the "sensory innovation," is then optimally scaled by the Kalman gain, K_k, to produce an a posteriori estimate of the state of the arm:

$$\hat{x}_k = \hat{x}_k + K_k\left(y_k - H\hat{x}_k^-\right) \tag{3}$$

In brief, discrete state estimation consists of a two-step recursive procedure such that the forward model generates an a priori estimate of the state, which is next refined by potentially innovative information gleaned from the sensory input to form the final, a posteriori estimate. PPC, specifically the parietal reach region (PRR) and area 5, seems to be a reasonable site for an observer for on-line control to reside, since it has access to two key inputs to the observer model: a large number of internal feedback connections from frontal areas (i.e., efference copy) and substantial sensory input from both visual and somatosensory domains (Johnson et al., 1996; Jones & Powell, 1970).

Continuous sensorimotor control and state estimation in PPC

Clinical and psychophysical studies in humans have established that PPC is involved not only in specifying movement plans, but also in the execution and control of ongoing movement. For example, it is well known that lesions in parietal cortex often lead to optic ataxia, that is, impairment in locating and reaching to stimuli in three-dimensional space (Balint, 1909; Perenin & Vighetto, 1988; Rondot, Recondo, & Ribadeaudumas, 1977). For instance, optic ataxia patients have difficulty making rapid and "automatic" corrective movements when guiding the hand to targets that have been jumped (Pisella et al., 2000). Similarly, Grea and colleagues (2002) reported a patient with bilateral parietal lesions who was unable to amend her movement to pick up a cylinder after it had been jumped to a new location at movement onset. Interestingly, instead of making corrective movements during an initial trajectory, the subject needed to perform two distinct movements: one that represented the initial plan and a second movement to reach to the new location of the cylinder. Using TMS applied to the posterior parietal cortex, Desmurget and colleagues (1999) were able to transiently disrupt the ability of most of their subjects to correct reaching trajectories made to targets that were displaced around the time of movement onset. Later, Della-Maggiore, Malfait, Ostry, and Pans (2004) showed that TMS applied to PPC interfered with the ability of subjects to adapt to novel force-field environments. An intriguing, potentially unifying explanation for all of these deficits, which was originally suggested by Wolpert, Goodbody, and Husain (1998), is that PPC may serve as an observer, which forms an internal estimate of the state of the arm during movement. A failure to accurately maintain this estimate on-line could result in an inability to monitor and therefore correct an ongoing movement. For example, Wolpert, Goodbody, and Husain reported a parietal lesion patient who was unable to maintain an internal estimate of the state of her hand. She could not maintain a constant precision grip force in absence of vision; with no vision of her stationary arm, she perceived it to drift slowly in space over 10–20 seconds until eventually reporting it to disappear. When she was asked to make slow-pointing movements to peripheral targets while maintaining central fixation, large errors accumulated in her trajectories (although self-paced movements were not impaired).

Mental Simulation of Movement Evidence that PPC is involved in sensorimotor state estimation also comes from the study of the mental simulation of movement, which presumably activates circuits that overlap with those engaged during motor control but inhibits execution of a movement itself (Decety, 1996; Gerardin et al., 2000; Stephan et al., 1995). When normal healthy subjects imagine making a goal-directed movement, mental simulation time typically matches the time needed to execute that same movement (Decety & Michel, 1989; Donders, 1969). This suggests that the brain is able to maintain a realistic estimate of the state of the hand over time while imagining a movement, despite sensory feedback being unavailable. Shadmehr and Krakauer (2008) interpreted this finding in the context of observer theory, suggesting that this capability indicates that the brain/observer is able to rely entirely upon the output of a forward model (in the absence of sensory feedback) to estimate the state of the arm during mental simulation (e.g., Kalman gain in equation 3 is set to zero). Interestingly, patients with unilateral motor cortex lesions (Sirigu et al., 1995) who show prolonged movement times compared to normal control subjects are still able to accurately imagine the duration of their movements (i.e., the simulation time and execution time remain well matched for these patients). Therefore, aberrant motor commands (u in equation 1) that are produced by the motor cortex could theoretically still be used by an intact observer to predict the correct temporal sequence of hand states (and therefore the trajectory duration), even for an impaired movement. Similarly, patients with lesions of the cerebellum (Kagerer, Bracha, Wunderlich, Stelmach, & Bloedel, 1998) and of the basal ganglia (Dominey, Decety, Broussolle, Chazot, & Jeannerod, 1995) also do not show a difference between simulation and execution times.

While M1, the cerebellum, and the basal ganglia do not appear to be critically involved in state estimation during simulated movements, PPC, by contrast, does appear to be essential for maintaining an internal representation of the state of the hand, which is necessary for producing a consistent relationship between simulation and execution time. Sirigu and colleagues (1996) later reported an impairment in the ability to simulate a movement in patients with right PPC lesions: the time needed to mentally simulate a movement was significantly different (generally less) than the time needed to execute the same movement. (Note that, similar to motor cortex lesion patients, actual execution time was

prolonged in comparison to control subjects.) This inconsistency suggests that the brain was unable to reliably estimate the state of the hand after damage to PPC. This impairment could be explained by multiple possible failures of the observer model: (1) an error in the forward model (i.e., faulty A or B matrices in equation 1), (2) an error when incorporating sensory feedback into the a priori estimate of the forward model (i.e., faulty H or K matrices in equation 3) or (3) a combination of these. On the basis of known strong sensory input to PPC, it is probable that PPC is involved in integrating sensory feedback into the state estimate. However, because visual and proprioceptive inputs were effectively removed during the above mental simulation tasks (e.g., eyes were closed, muscle activity was absent), it is less likely that erroneous state estimation was due exclusively to faulty integration of sensory feedback. Also, most parietal lesion patients significantly underestimated the time it would take to complete a movement when simulating it. Such a systematic decrease in imagined movement duration may have arisen due to an erroneous a priori estimate made by a forward model, whose transition matrices A and B govern the rate at which the arm propagates through space. Therefore these mental simulation results suggest that PPC is also involved in propagating the state of the arm forward in time using a forward model (equation 1). If we assume that PPC incorporates sensory information into the forward model state estimate as well, then PPC would be best described as an observer, as Wolpert and colleagues suggested.

Neural correlates of sensorimotor state estimation in PPC

Psychophysical and clinical reports have pointed to both the parietal lobe and the cerebellum as candidate neural substrates for a forward model (Blakemore & Sirigu, 2003; Miall et al., 1993; Wolpert, Goodbody, & Husain, 1998; Wolpert, Miall, & Kawato, 1998). For example, Desmurget and colleagues suggested that PPC encodes a forward model of the arm's dynamics, from which it may also compute an estimate of the motor error (i.e., the difference between the target vector and the movement vector), which could then be transformed into a corrective motor command by the cerebellum (Desmurget & Grafton, 2000). While numerous studies have shown that PPC and the cerebellum are likely to be involved in forward model control, finding direct neural correlates of forward model state estimation in the brain has proven difficult.

On-line Directional Control Signals in PPC Previous encoding studies have shown that area 5 neurons are correlated with a variety of movement- and task-related parameters (most notably velocity and target position) during reaching movements made with a manipulandum (Ashe & Georgopoulos, 1994; Averbeck, Chafee, Crowe, & Georgopoulous, 2005). These studies concluded that area 5 largely encodes a sensory (i.e., proprioceptive) representation that slightly lags the state of the movement (i.e., lag time = −30 ms). More recently, we further investigated the neural representation of on-line directional control signals in both area 5 and MIP while monkeys performed center-out and obstacle avoidance joystick trajectories under central eye fixation (Mulliken, Musallam, & Andersen, 2008a) (figures 41.2*A* and 41.2*B*). We analyzed the correlations of single neurons recorded in both area 5 and PRR with the static goal angle (fixed angle from the starting cursor position to the target) and the dynamic movement angle of the cursor (angle of heading) during a joystick task.

To characterize a neuron's dynamic tuning for movement angle, we constructed a space-time tuning function (STTF). Each horizontal slice in the STTF plots a neuron's instantaneous firing rate as a function of the angle measured at a particular lag time (Paninski, Fellows, Hatsopoulos, & Donoghue, 2004). For each lag time in the STTF, we also calculated the mutual information between firing rate and movement angle. The resultant temporal encoding function (TEF) indicated how strongly a neuron's instantaneous firing rate encoded the movement angle at different lag times (i.e., from past (lag time < 0) to future (lag time > 0) angles). The lag time corresponding to the peak of the TEF was considered to be the optimal lag time (OLT). Figure 41.2*C* shows a representative movement angle STTF for a single neuron. This neuron encoded the most information about the movement angle at an OLT of 0 ms and therefore best encoded the current state of the movement angle (figure 41.2*D*).

For our PPC population, during the center-out task, 56% of task-related neurons encoded significant information about the movement angle, and 75% of these significantly encoded the goal angle (note that PPC neurons appeared to be more engaged during the obstacle task: 79% encoded movement angle, and 93% encoded goal angle). Interestingly, we found an anatomical correlate for the representation of goal angle and movement angle in PPC: Mutual information for goal angle increased gradually with recording depth in the sulcus, while movement angle information (peak information measured at OLT) decreased with depth. A stronger encoding of target-related signals deeper in the intraparietal sulcus (IPS) and, conversely, a favored representation of hand movement–related activity in surface regions of the IPS are consistent with findings from previous PPC studies of reach planning, in which eye-centered target signals were commonly found in deeper structures such as PRR and more hand-related activity was reported for surface area 5 neurons (Buneo et al., 2002).

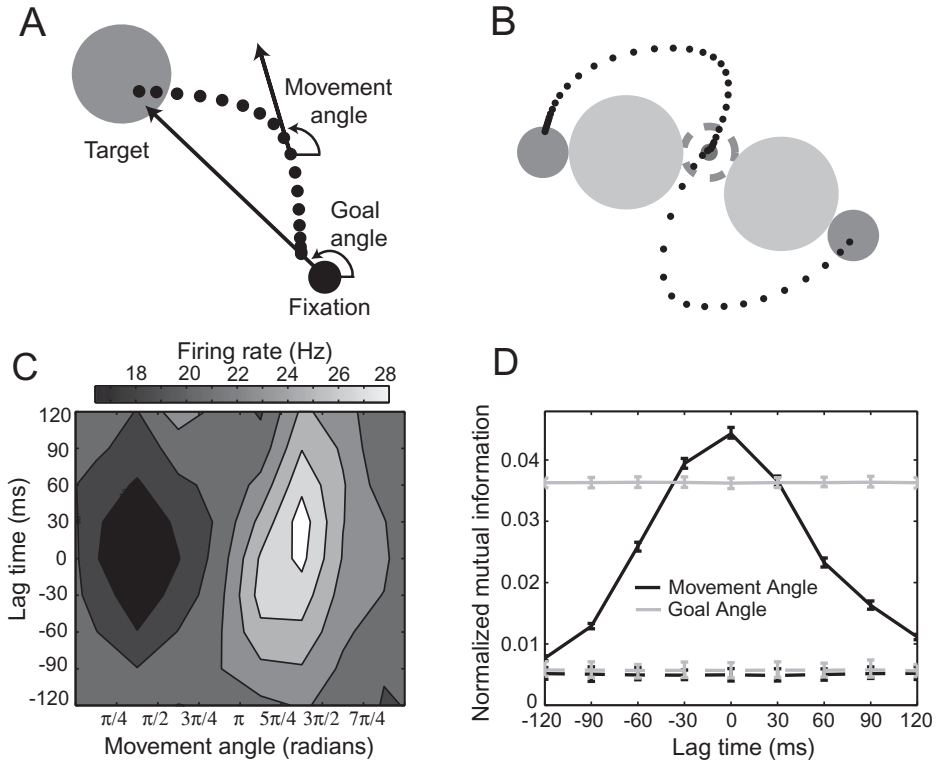


FIGURE 41.2 Experimental design and representative neuron. (*A*) Example center-out trajectory showing the goal angle and movement angle, and their respective origins of reference. Large and medium-sized circles represent the target and fixation point, respectively. Dots denote cursor position sampled at 15-ms intervals along the trajectory. (*B*) Example trajectories for obstacle task. The dashed circle depicts the starting location of the target and is not visible once the target has been jumped to the periphery. The large gray circles represent the visual obstacle. (*C*) Movement angle space-time tuning function (STTF). The contour plot shows the average firing rate of a cell that occurred for different movement angles measured over a range of lag times (-120 ms $\leq \tau \leq 120$ ms) relative to the firing rate. (*D*) Movement angle temporal encoding function (TEF) and corresponding goal angle TEF, where mutual information between firing rate and movement angle is plotted as a function of lag time. The firing rate contained the most information about the movement angle at an optimal lag time of 0 ms. The dashed lines denote surrogate TEFs, for both movement (black-dashed) and goal (gray-dashed) angles, that were derived from surrogate spike trains and actual angles. (Reprinted with permission from Mulliken, Musallam, & Andersen, 2008a.)

Neurons that are significantly tuned for goal angle persistently encode the static direction to the target, independent of the changing state of the cursor. These cells were consistent with previous reports of target-sensitive tuning in area 5 (Ashe & Georgopoulos, 1994). Therefore, the intended goal of the trajectory is maintained in PPC during on-line control of movement. PPC neurons that are tuned for movement angle encode dynamic information about the time-varying state of the cursor. Figure 41.3*A* shows TEFs for the movement angle population. The histogram in figure 41.3*B* summarizes the distribution of OLTs for the movement angle population, which was centered at 0 ± 90 ms and 30 ± 90 ms, for the center-out and obstacle tasks, respectively (median ± interquartile range (IQR)). These plots show that movement angle neurons contained a temporal distribution of information about the state of the ongoing movement; some neurons best represented states in the near future (positive-lag time), some best represented states in the recent past (negative-lag time), and many peaked around the current state (zero-lag time).

It is helpful to interpret the OLT results in the context of the observer framework. Passive sensory feedback (e.g., y in equation 2) would require at least 30–90 ms (proprioceptive-visual) to reach PPC; consistent with some of the negative OLTs (≤–30 ms) observed here (Decety et al., 1994; Flanders & Cordo, 1989; Miall & Wolpert, 1996; Petersen et al., 1998; Raiguel et al., 1999). Conversely, if PPC neurons were responsible for generating outgoing motor commands (u in equation 1), subsequent stages of processing and execution of the movement would require at least 90–100 ms to produce the corresponding cursor motion (Miall & Wolpert, 1996). For instance, similar analyses for velocity have been performed in the primary motor cortex and report average OLTs of approximately 90–100 ms (Ashe & Georgopoulos, 1994; Paninski et al., 2004). Therefore, it is unlikely that PPC is primarily driving motor

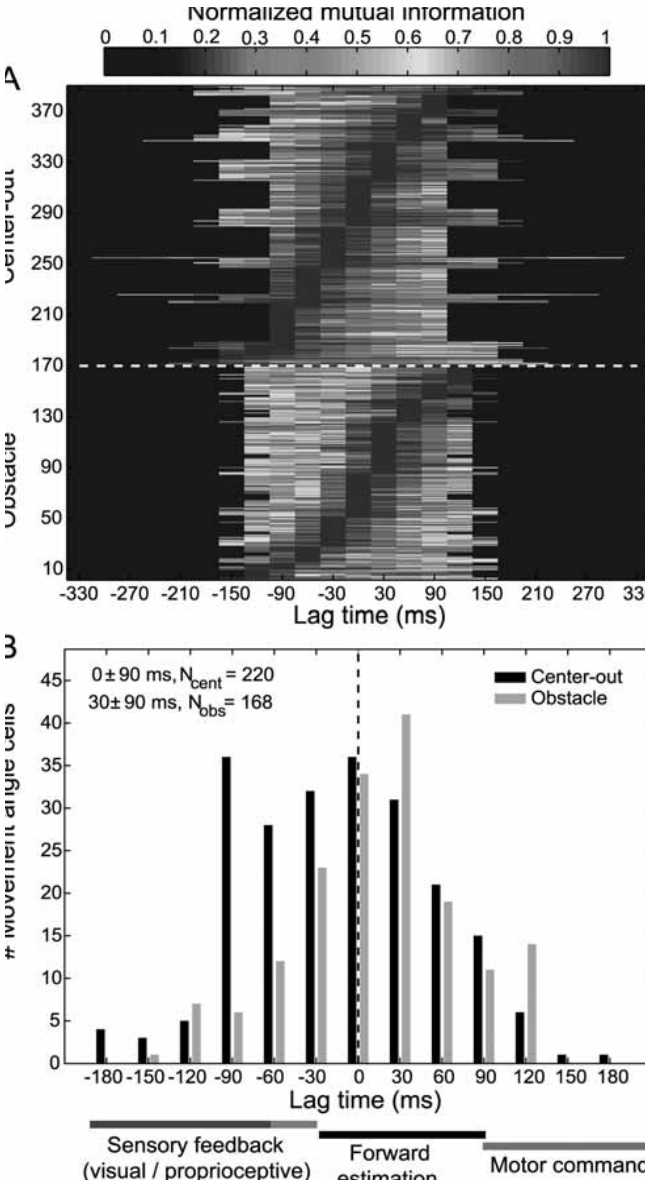


FIGURE 41.3 Population temporal encoding results. (*A*) Population TEFs plotted for all movement angle neurons showing cell-normalized mutual information as a function of lag time. (*B*) Histogram summarizing the OLTs for movement angle neurons for both center-out and obstacle tasks (summary statistic in upper-left corner: median ± interquartile range). Many of these neuron's OLTs were consistent with a forward estimate of the state of the movement angle, which did not directly reflect delayed sensory feedback to PPC, nor were they compatible with outgoing motor commands from PPC. (Reprinted with permission from Mulliken, Musallam, & Andersen, 2008a.) (See color plate 53.)

cortex with feedforward commands, since it would be expected that PPC should lead the movement state by more than motor cortex does, on average (i.e., OLT > 90 ms). Neither passive sensory feedback nor efferent motor explanations best account for the responses of neurons whose OLTs fall between −30 and 60 ms. Instead, these cells appear to encode a forward-state estimate, which allows PPC to monitor the current and upcoming states of the movement angle prior to the arrival of delayed sensory feedback. It does not appear that this current-state estimate is merely a blend of incoming sensory and outgoing motor representations (i.e., a simple summation of two modal distributions centered at negative and positive lag times should result in a bimodal or potentially "flat" distribution), since our OLT distributions appear to show a pronounced unimodal peak around 0 ms. Furthermore, the peak information (mutual information at the OLT) encoded by neurons that were most clearly forward-estimating ($0 \leq \text{OLT} \leq 60$ ms) was significantly larger than the peak information encoded by the remaining population of movement angle neurons (OLT ≤ −30 ms, or OLT ≥ 90 ms). Therefore not only does PPC have a central tendency to encode the current state of the movement angle, but forward-estimating neurons also contained significantly more information about the movement state than did neurons with other OLTs, suggesting that these state estimates are generated by some active computational process (i.e., a forward model).

While it is likely that PPC relies upon a forward model to estimate the current state of the cursor (i.e., a priori estimate), it is also possible that sensory information is integrated by PPC to update this estimate. As mentioned above, it has been suggested that the a priori state estimate is generated by the cerebellum and then sent to PPC (Shadmehr & Krakauer, 2008). In this situation, these authors suggested that PPC is responsible only for processing afferent signals (i.e., matrix H in equation 2), specifying the Kalman gain to optimally incorporate sensory information into a refined, a posteriori state estimate. Given known afferent projections to PPC (both visual and proprioceptive) as well as evidence from our data demonstrating that some movement angle neurons appear to encode a passive sensory representation of the state (i.e., OLT <= −30 ms), it seems likely that PPC does integrate delayed sensory information. However, on the basis of our data and evidence from the mental simulation literature (discussed above), we suggest that PPC is also involved directly in performing forward model computations, perhaps within a reciprocal, functional loop that includes the cerebellum (Blakemore & Sirigu, 2003). That is, the forward state estimates found in PPC most likely reflect the output of an observer, which is involved in both performing the computations of the forward model and integrating sensory feedback into the state estimate.

DYNAMIC TUNING AND SEPARABILITY OF MOVEMENT ANGLE STTF Further support for state estimation in PPC was obtained from analyzing the spatiotemporal encoding properties of movement angle STTFs. We measured changes in the preferred direction of a neuron, θ_{pd}, over a range of

lag times. θ_{pd} is the movement angle at which a neuron fired maximally for a particular lag time. We reasoned that if θ_{pd} did not vary significantly as a function of lag time compared to changes that occurred in the movement angle itself, then that neuron encoded a mostly straight trajectory. Across the population of movement angle neurons, most neurons' STTFs exhibited small changes in θ_{pd} as a function of lag time, which were significantly less than changes observed in the actual movement angle in the trajectories themselves (figure 41.4*AB*).

We performed an additional separability analysis to further characterize the relationship between angle and lag time encoded by a neuron's STTF. A perfectly separable STTF indicates that the lag time and angle were encoded independently of one another. We determined that the population of movement angle neurons was largely separable in the angle-time plane by using singular value decomposition (SVD) (Mazer, Vinje, McDermott, Schiller, & Gallant, 2002; Pena & Konishi, 2001). We calculated the fractional energy contained in the singular values for each cell's movement angle STTF; 92.0 ± 14.7% and 78.9 ± 25.8% of energy (median + IQR) was contained in the first singular value, for the center-out and obstacle tasks, respectively (figure 41.4*C*). The distribution of fractional energies contained in the first singular value is shown in figure 41.4*D*.

These results suggest that dynamic sensorimotor control mechanisms in PPC encode mostly straight and instantaneous trajectories, with a less substantial component of the neurons' firing rates arising because of nonlinear encoding mechanisms that may reflect the slight curvature we observed in the STTFs. This interpretation is consistent with PPC neurons encoding a state estimate of the movement direction, such that the majority of information is encoded at a cell's OLT, with decreasing information encoded away from the OLT. (Note that a perfectly instantaneous state estimate, that is, a delta function, should not be expected due to autocorrelation present in continuous motor variables such as movement angle.)

Reading out the dynamic state of a cursor from PPC

It would be interesting to test whether a dynamic state estimate in PPC, presumably reflecting the operation of an observer, could be used to causally control an external device

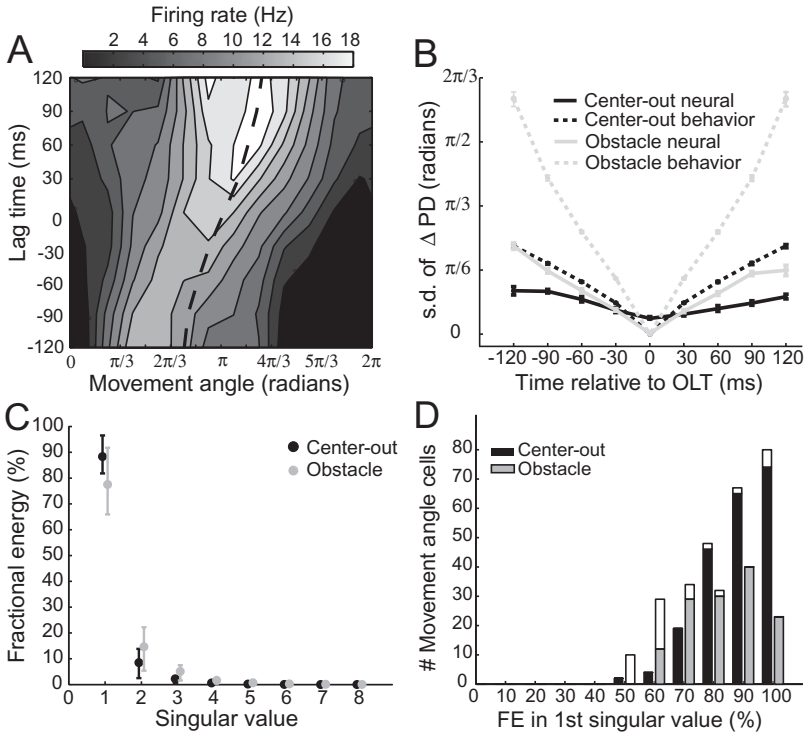


Figure 41.4 Curvature and separability of STTFs. (*A*) Example STTF containing slight curvature. The θ_{pd} of this cell (dashed line) changed smoothly but slightly as a function of lag time. (*B*) Standard deviation of the population's distribution of θ_{pd} changes ($\sigma_{d\theta}$), plotted as a function of time relative to the OLT. For both center-out and obstacle tasks, the population $\sigma_{d\theta}$ (neural, solid lines) was significantly less than the $\sigma_{d\theta}$ for the actual movement angle (behavior, dashed lines) over the same time range. (*C*) Population summary of fractional energy (FE) accounted for by each singular vector in the singular vector decomposition (SVD) analysis. The majority of energy in movement angles STTFs was captured by the first singular vectors for the center-out and obstacle tasks, respectively. (*D*) Population histogram showing distribution of FE of the first singular value for all movement angle cells. (Reprinted with permission from Mulliken, Musallam, & Andersen, 2008a.)

besides our own limbs. During recent years, several groups have leveraged the findings from decades of primate neurophysiology toward the development of an important medical application: a neural prosthesis to assist paralyzed individuals. A neural prosthesis would directly read out the desired movement intentions of a patient from regions of the brain that are not affected by injury or disease.

Several groups have successfully extracted continuous movement information (i.e., trajectories) from motor cortices, such as M1 and dorsal premotor cortex (PMd) (Carmena et al., 2003; Kennedy, Bakay, Moore, Adams, & Goldwaithe, 2000; Musallam et al., 2004; Patil, Carmena, Nicolelis, & Turner, 2004; Santhanam, Ryu, Yu, Afshar, & Shenoy, 2006; Serruya, Hatsopoulos, Paninski, Fellows, & Donoghue, 2002; Shenoy et al., 2003; Taylor, Tillery, & Schwartz, 2002; Wessberg et al., 2000; Wolpaw & McFarland, 2004). In contrast to signals extracted from M1, which are more likely to encode movement execution signals that are represented in a musculoskeletal reference frame, high-level visuomotor signals can be found in earlier stages of the dorsal visual pathway, such as in PPC or PMd. For example, the goal of a reach in visual coordinates has been decoded successfully from both PPC and PMd neurons (Musallam, Corneil, Greger, Scherberger, & Andersen, 2004; Santhanam et al., 2006). Sensorimotor areas of cortex, particularly those that are strongly innervated by visual feedback projections (e.g., PPC) represent candidate regions that are potentially useful for driving a neural prosthesis since a primary source of input, visual information, is typically uncompromised after paralysis (figure 41.5).

Offline Decoding of Trajectories We recently built upon the work of Musallam and colleagues and demonstrated that a PPC prosthesis can also be used to perform continuous control of a computer cursor (Mulliken, Musallam, & Anderson, 2008b). First, we showed that we could reliably reconstruct monkeys' trajectories off-line using a small ensemble of PPC cells. For example, decoding from just five single neurons using a Kalman filter, we demonstrated that we could account for more than 70% of the variance in the cursor position. Interestingly, by extracting information about the goal of a trajectory (i.e., target information that is also known to be encoded in PPC) and incorporating it into the Kalman filter framework, we were able to significantly improve the accuracy of the decoded estimate (on average by 17% over a standard Kalman filter).

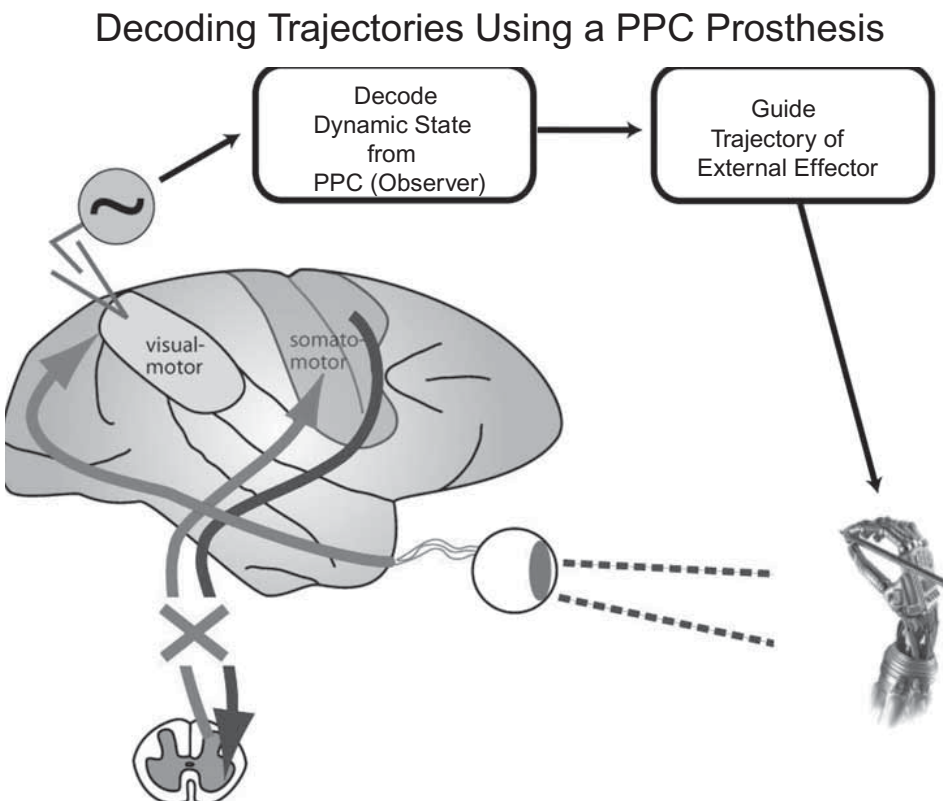


Figure 41.5 A neural prosthesis using PPC for trajectory control. A spinal cord injury can render communication (afferent and efferent) between somatosensory and motor areas of cortex and the limbs useless. However, the integrity of the "vision for action" pathway may still be largely intact, which includes PPC. Decoding algorithms are designed to optimally estimate the state of the effector from the measurement of neural activity from PPC ensembles.

In these decoding experiments, we presumably were decoding from the output of an observer, thereby harnessing a forward estimate of the expected (e.g., current) sensorimotor state of the cursor. To verify this, we also decoded the state (in this situation, the position and velocity) of the cursor shifted in time relative to the instantaneous firing rate measurement, with lag times ranging from −300 ms to 300 ms, in 30-ms steps (where negative lag times correspond to past movement states and positive lag times correspond to future movement states). The optimal lag time (OLT) for decoding velocity using the G-Kalman filter was 10 ms in the future, consistent with previous claims that PPC best represents the current state of the velocity (Mulliken, Musallam, & Andersen, 2008a). The position of the cursor was best decoded slightly further into the future, at an OLT of approximately 40 ms. These temporal decoding results suggest that the current or upcoming state of the cursor could be best extracted from the PPC population by using the Kalman filter. These results are by and large similar to the encoding analyses reported above, and they suggest that PPC is involved in maintaining an estimate of the current and upcoming state of the cursor, consistent with the output of a forward model.

Closed Loop Brain Control Decoding In addition to off-line decoding, we demonstrated that we could decode trajectories during closed loop brain control sessions, in which the real-time position of the cursor was determined solely by a monkey's thoughts. Initially, the monkey performed brain trajectories at approximately a 30% success rate (for eight targets), but he quickly improved his performance to an 80% success rate after just four to five sessions. This increase in behavioral performance was accompanied by a corresponding enhancement in neural tuning properties, showing that learning effects occurred in the PPC ensemble. For instance, off-line analyses showed that the neurons' average tuning depth increased by more than 70%, the average coverage of two-dimensional space of the population increased by 35%, and the off-line decoding performance (i.e., R^2) of the PPC ensemble increased by more than twofold. These data show that PPC ensembles can be harnessed independently for real-time continuous control of a cursor. In addition, the ability of PPC to causally control a cursor indicates that the state representation in PPC does not rely entirely on visual/proprioceptive information but instead may reflect current and future state estimates generated by a forward model. Last, we expect, on the basis of our findings here and PPC's known functional role in combining visual and motor representations, that PPC will be particularly well suited to serve as a target for a prosthesis that relies upon visually guided feedback for continuous control and error-driven learning.

The ability to extract both trajectory and goal information from neural activity makes this brain area an attractive target for a neural prosthesis. For example, a continuous decoder that estimates the dynamic state of the cursor could be improved by using target information to constrain the decoded trajectory on the basis of its inferred endpoint (Srinivasan & Brown, 2007). The observation that these neurons appear to encode mostly straight lines in visual space may prove to be more straightforward to decode. For instance, PPC neurons may be more flexible for controlling a variety of end effectors, including but not limited to the human arm. Finally, when training a prosthetic in a clinical setting, the operator must rely on a patient's ability to imagine moving an effector in space. Motor imagery studies suggest that PPC is a critical node for maintaining an accurate estimate of the state of the hand during mental simulation of movement. Therefore, we expect that PPC will be a useful site for extracting time-varying trajectory information that accurately matches the desired, real-time sensory outcome of an intended movement trajectory. These findings mark an important step forward in the development of a neural prosthesis using signals from PPC.

REFERENCES

Andersen, R. A., & Buneo, C. A. (2002). Intentional maps in posterior parietal cortex. *Annu. Rev. Neurosci.*, *25*, 189–220.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1987). Neurons of area-7 activated by both visual stimuli and oculomotor behavior. *Exp. Brain Res.*, *67*(2), 316–322.

Ashe, J., & Georgopoulos, A. P. (1994). Movement parameters and neural activity in motor cortex and area 5. *Cereb. Cortex*, *4*(6), 590–600.

Atkeson, C. G. (1989). Learning arm kinematics and dynamics. *Annu. Rev. Neurosci.*, *12*, 157–183.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2005). Parietal representation of hand velocity in a copy task. *J. Neurophysiol.*, *93*(1), 508–518.

Balint, R. (1909). Seelenlahmung des "Schauens," optische Ataxie, raumliche Storung der Aufmerksamkeit. *Monatsschr. Psychiatr. Neurol.*, *25*, 51–81.

Batista, A. P., Buneo, C. A., Snyder, L. H., & Andersen, R. A. (1999). Reach plans in eye-centered coordinates. *Science*, *285*(5425), 257–260.

Blakemore, S. J., & Sirigu, A. (2003). Action prediction in the cerebellum and in the parietal lobe. *Exp. Brain Res.*, *153*(2), 239–245.

Bradley, D. C., Maxwell, M., Andersen, R. A., Banks, M. S., & Shenoy, K. V. (1996). Mechanisms of heading perception in primate visual cortex. *Science*, *273*(5281), 1544–1547.

Buneo, C. A., Jarvis, M. R., Batista, A. P., & Andersen, R. A. (2002). Direct visuomotor transformations for reaching. *Nature*, *416*(6881), 632–636.

Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., et al. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.*, *1*(2), 193–208.

Claxton, G. (1975). Why can't we tickle ourselves? *Percept. Mot. Skills*, *41*(1), 335–338.

Connolly, J. D., Andersen, R. A., & Goodale, M. A. (2003). FMRI evidence for a 'parietal reach region' in the human brain. *Exp. Brain Res.*, *153*(2), 140–145.

Cullen, K. E. (2004). Sensory signals during active versus passive movement. *Curr. Opin. Neurobiol.*, *14*(6), 698–706.

Decety, J. (1996). Do imagined and executed actions share the same neural substrate? *Cogn. Brain Res.*, *3*(2), 87–93.

Decety, J., & Michel, F. (1989). Comparative analysis of actual and mental movement times in 2 graphic tasks. *Brain Cogn.*, *11*(1), 87–97.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., et al. (1994). Mapping motor representations with positron emission tomography. *Nature*, *371*(6498), 600–602.

Della-Maggiore, V., Malfait, N., Ostry, D. J., & Paus, T. (2004). Stimulation of the posterior parietal cortex interferes with arm trajectory adjustments during the learning of new dynamics. *J. Neurosci.*, *24*(44), 9971–9976.

Desmurget, M., Epstein, C. M., Turner, R. S., Prablanc, C., Alexander, G. E., & Grafton, S. T. (1999). Role of the posterior parietal cortex in updating reaching movements to a visual target. *Nat. Neurosci.*, *2*(6), 563–567.

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends Cogn. Sci.*, *4*(11), 423–431.

DeSouza, J. F. X., Dukelow, S. P., Gati, J. S., Menon, R. S., Andersen, R. A., & Vilis, T. (2000). Eye position signal modulates a human parietal pointing region during memory-guided movements. *J. Neurosci.*, *20*(15), 5835–5840.

Dominey, P., Decety, J., Broussolle, E., Chazot, G., & Jeannerod, M. (1995). Motor imagery of a lateralized sequential task is asymmetrically slowed in hemi-Parkinsons patients. *Neuropsychologia*, *33*(6), 727–741.

Donders, F. C. (1969). On speed of mental processes. *Acta Psychol. (Amst.)*, *30*, 412–431.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye-movements. *Science*, *255*(5040), 90–92.

Eskandar, E. N., & Assad, J. A. (1999). Dissociation of visual, motor and predictive signals in parietal cortex during visual guidance. *Nat. Neurosci.*, *2*(1), 88–93.

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, A. (2003). Modulating the experience of agency: A positron emission tomography study. *NeuroImage*, *18*(2), 324–333.

Fink, G. R., Marshall, J. C., Halligan, P. W., Frith, C. D., Driver, J., Frackowiak, R. S. J., et al. (1999). The neural consequences of conflict between intention and the senses. *Brain*, *122*, 497–512.

Flanders, M., & Cordo, P. J. (1989). Kinesthetic and visual control of a bimanual task: Specification of direction and amplitude. *J. Neurosci.*, *9*(2), 447–453.

Gail, A., & Andersen, R. A. (2006). Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations. *J. Neurosci.*, *26*(37), 9376–9384.

Georgopoulos, A. P., Kalaska, J. F., & Massey, J. T. (1981). Spatial trajectories and reaction-times of aimed movements: Effects of practice, uncertainty, and change in target location. *J. Neurophysiol.*, *46*(4), 725–743.

Gerardin, E., Sirigu, A., Lehericy, S., Poline, J. B., Gaymard, B., Marsault, C., et al. (2000). Partially overlapping neural networks for real and imagined hand movements. *Cereb. Cortex*, *10*(11), 1093–1104.

Geshwind, N., & Damasio, A. R. (1985). Apraxia. In P. J. Vinken, G. W. Bruyn, & H. L. Klawans (Eds.), *Handbook of clinical neurology* (pp. 423–432). Amsterdam: Elsevier.

Goodwin, G. C., & Sin, K. S. (1984). *Adaptive filtering prediction and control.* Englewood Cliffs, NJ: Prentice-Hall.

Grea, H., Pisella, L., Rossetti, Y., Desmurget, M., Tilikete, C., Grafton, S., et al. (2002). A lesion of the posterior parietal cortex disrupts on-line adjustments during aiming movements. *Neuropsychologia*, *40*(13), 2471–2480.

Haarmeier, T., Bunjes, F., Lindner, A., Berret, E., & Thier, P. (2001). Optimizing visual motion perception during eye movements. *Neuron*, *32*(3), 527–535.

Haarmeier, T., Thier, P., Repnow, M., & Petersen, D. (1997). False perception of motion in a patient who cannot compensate for eye movements. *Nature*, *389*(6653), 849–852.

Johnson, P. B., Ferraina, S., Bianchi, L., & Caminiti, R. (1996). Cortical networks for visual reaching: Physiological and anatomical organization of frontal and parietal lobe arm regions. *Cereb. Cortex*, *6*(2), 102–119.

Jones, E. G., & Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, *93*(4), 793–820.

Jordan, M. I. (1995). Computational motor control. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 597–609). Cambridge, MA: MIT Press.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognit. Sci.*, *16*(3), 307–354.

Kagerer, F. A., Bracha, V., Wunderlich, D. A., Stelmach, G. E., & Bloedel, J. R. (1998). Ataxia reflected in the simulated movements of patients with cerebellar lesions. *Exp. Brain Res.*, *121*(2), 125–134.

Kalaska, J. F., Caminiti, R., & Georgopoulos, A. P. (1983). Cortical mechanisms related to the direction of two-dimensional arm movements: Relations in parietal area 5 and comparison with motor cortex. *Exp. Brain Res.*, *51*(2), 247–260.

Kalaska, J. F., & Crammond, D. J. (1995). Deciding not to go: Neuronal correlates of response selection in a go/nogo task in primate premotor and parietal cortex. *Cereb. Cortex*, *5*(5), 410–428.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME-J. Basic Eng.*, *82*, 35–45.

Kawato, M., Furukawa, K., & Suzuki, R. (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biol. Cybern.*, *57*(3), 169–185.

Kennedy, P. R., Bakay, R. A. E., Moore, M. M., Adams, K., & Goldwaithe, J. (2000). Direct control of a computer from the human central nervous system. *IEEE Trans. Rehab. Eng.*, *8*(2), 198–202.

MacDonald, P. A., & Paus, T. (2003). The role of parietal cortex in awareness of self-generated movements: A transcranial magnetic stimulation study. *Cereb. Cortex*, *13*(9), 962–967.

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proc. Natl. Acad. Sci. USA*, *99*(3), 1645–1650.

Miall, R. C., Weir, D. J., Wolpert, D. M., & Stein, J. F. (1993). Is the cerebellum a Smith predictor? *J. Motor Behav.*, *25*(3), 203–216.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Net.*, *9*(8), 1265–1279.

Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H., & Acuna, C. (1975). Posterior parietal association cortex of

monkey: Command functions for operations within extrapersonal space. *J. Neurophysiol.*, *38*(4), 871–908.

Mulliken, G. H., Musallam, S., & Andersen, R. A. (2008a). Forward estimation of movement state in posterior parietal cortex. *Proc. Natl. Acad. Sci. USA*, *105*(24), 8170–8177.

Mulliken, G. H., Musallam, S., & Andersen, R. A. (2008b). *Decoding trajectories from posterior parietal cortex. J. Neurosci.*, *28*(48), 12913–12926.

Musallam, S., Corneil, B. D., Greger, B., Scherberger, H., & Andersen, R. A. (2004). Cognitive control signals for neural prosthetics. *Science*, *305*(5681), 258–262.

Paninski, L., Fellows, M. R., Hatsopoulos, N. G., & Donoghue, J. P. (2004). Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *J. Neurophysiol.*, *91*(1), 515–532.

Patil, P. G., Carmena, L. M., Nicolelis, M. A. L., & Turner, D. A. (2004). Ensemble recordings of human subcortical neurons as a source of motor control signals for a brain-machine interface. *Neurosurgery*, *55*(1), 27–35.

Pellijeff, A., Bonilha, L., Morgan, P. S., McKenzie, K., & Jackson, S. R. (2006). Parietal updating of limb posture: An event-related fMRI study. *Neuropsychologia*, *44*(13), 2685–2690.

Pena, J. L., & Konishi, M. (2001). Auditory spatial receptive fields created by multiplication. *Science*, *292*(5515), 249–252.

Perenin, M. T., & Vighetto, A. (1988). Optic ataxia: A specific disruption in visuomotor mechanisms: 1. Different aspects of the deficit in reaching for objects. *Brain*, *111*, 643–674.

Petersen, N., Christensen, L. O. D., Morita, H., Sinkjaer, T., & Nielsen, J. (1998). Evidence that a transcortical pathway contributes to stretch reflexes in the tibialis anterior muscle in man. *J. Physiol.*, *512*(1), 267–276.

Pisella, L., Grea, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., et al. (2000). An 'automatic pilot' for the hand in human posterior parietal cortex: Toward reinterpreting optic ataxia. *Nat. Neurosci.*, *3*(7), 729–736.

Poulet, J. F. A., & Hedwig, B. (2003). Corollary discharge inhibition of ascending auditory neurons in the stridulating cricket. *J. Neurosci.*, *23*(11), 4717–4725.

Raiguel, S. E., Xiao, D. K., Marcar, V. L., & Orban, G. A. (1999). Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. *J. Neurophysiol.*, *82*(4), 1944–1956.

Robinson, D. L., Goldberg, M. E., & Stanton, G. B. (1978). Parietal association cortex in primate: Sensory mechanisms and behavioral modulations. *J. Neurophysiol.*, *41*(4), 910–932.

Rondot, P., Recondo, J. D., & Ribadeaudumas, J. L. (1977). Visuomotor ataxia. *Brain*, *100*(June), 355–376.

Roy, J. E., & Cullen, K. E. (2004). Dissociating self-generated from passively applied head motion: Neural mechanisms in the vestibular nuclei. *J. Neurosci.*, *24*(9), 2102–2111.

Rushworth, M. F. S., Paus, T., & Sipila, P. K. (2001). Attention systems and the organization of the human parietal cortex. *J. Neurosci.*, *21*(14), 5262–5271.

Santhanam, G., Ryu, S. I., Yu, B. M., Afshar, A., & Shenoy, K. V. (2006). A high-performance brain-computer interface. *Nature*, *442*(7099), 195–198.

Seal, J., Gross, C., & Bioulac, B. (1982). Activity of neurons in area-5 during a simple arm movement in monkeys before and after deafferentation of the trained limb. *Brain Res.*, *250*(2), 229–243.

Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., & Donoghue, J. P. (2002). Instant neural control of a movement signal. *Nature*, *416*(6877), 141–142.

Shadmehr, R., & Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Exp. Brain Res.*, *185*(3), 359–381.

Shenoy, K. V., Meeker, D., Cao, S. Y., Kureshi, S. A., Pesaran, B., Buneo, C. A., et al. (2003). Neural prosthetic control signals from plan activity. *NeuroReport*, *14*(4), 591–596.

Sirigu, A., Cohen, L., Duhamel, J. R., Pillon, B., Dubois, B., Agid, Y., et al. (1995). Congruent unilateral impairments for real and imagined hand movements. *NeuroReport*, *6*(7), 997–1001.

Sirigu, A., Daprati, E., Pradat-Diehl, P., Franck, N., & Jeannerod, M. (1999). Perception of self-generated movement following left parietal lesion. *Brain*, *122*, 1867–1874.

Sirigu, A., Duhamel, J. R., Cohen, L., Pillon, B., Dubois, B., & Agid, Y. (1996). The mental representation of hand movements after parietal cortex damage. *Science*, *273*(5281), 1564–1568.

Snyder, L. H., Batista, A. P., & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature*, *386*(6621), 167–170.

Sommer, M. A., & Wurtz, R. H. (2006). Influence of the thalamus on spatial visual processing in frontal cortex. *Nature*, *444*(7117), 374–377.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.*, *43*(6), 482–489.

Srinivasan, L., & Brown, E. N. (2007). A state-space framework for movement control to dynamic goals through brain-driven interfaces. *IEEE Trans. Biomed. Eng.*, *54*(3), 526–535.

Stephan, K. M., Fink, G. R., Passingham, R. E., Silbersweig, D., Ceballosbaumann, A. O., Frith, C. D., et al. (1995). Functional anatomy of the mental representation of upper extremity movements in healthy subjects. *J. Neurophysiol.*, *73*(1), 373–386.

Taylor, D. M., Tillery, S. I. H., & Schwartz, A. B. (2002). Direct cortical control of 3D neuroprosthetic devices. *Science*, *296*(5574), 1829–1832.

Todorov, E. (2006). Optimal control theory. In K. Doya, S. Sihii, A. Pouget, & R. P. N. Rao (Eds.), *Bayesian brain: Probabilistic approaches to neural coding* (pp. 269–298). Cambridge, MA: MIT Press.

Wang, X. L., Zhang, M. S., Cohen, I. S., & Goldberg, M. E. (2007). The proprioceptive representation of eye position in monkey primary somatosensory cortex. *Nat. Neurosci.*, *10*(5), 640–646.

Weiskrantz, L., Elliott, J., & Darlington, C. (1971). Preliminary observations on tickling oneself. *Nature*, *230*(5296), 598–599.

Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., et al. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, *408*(6810), 361–365.

Wolpaw, J. R., & McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. USA*, *101*(51), 17849–17854.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882.

Wolpert, D. M., Goodbody, S. J., & Husain, M. (1998). Maintaining internal representations the role of the human superior parietal lobe. *Nat. Neurosci.*, *1*(6), 529–533.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci.*, *2*(9), 338–347.

42 Parallels between Sensory and Motor Information Processing

EMANUEL TODOROV

ABSTRACT The computational problems solved by the sensory and motor systems appear very different: One has to do with inferring the state of the world given sensory data, the other with generating motor commands appropriate for given task goals. However recent mathematical developments summarized in this chapter show that these two problems are in many ways related. Therefore information processing in the sensory and motor systems may be more similar than was previously thought—not only in terms of computations, but also in terms of algorithms and neural representations. Here, we explore these similarities and clarify some differences between the two systems.

Similarity between inference and control: An intuitive introduction

Consider a control problem in which we want to achieve a certain goal at some point in time in the future—say, grasp a coffee cup within 1 s. To achieve this goal, the motor system has to generate a sequence of muscle activations that result in joint torques that act on the musculoskeletal plant in such a way that the fingers end up curled around the cup. Actually, the motor system does not have to compute the entire sequence of muscle activations in advance. All it has to compute are the muscle activations right now, given the current state of the world (including the body) and some description of what the goal is. If the system is capable of performing this computation, then it will generate the resulting muscle activations, the clock will advance to the next point in time, and the computation will be repeated.

How can this control problem be interpreted as an inference problem? Instead of aiming for a goal in the future, imagine that the future is now and the goal has been achieved. More precisely, shift the time axis by 1 s and create a fictive sensory measurement corresponding to the hand grasping the cup. The inference problem is now as follows: Given that the fingers are around the cup and that the world was at a certain state 1 s ago, infer the muscle activations that caused the observed state transition. As in the control problem, all that needs to be inferred are the muscle activations at a single point in time (1 s ago); if this can be done, then the clock will advance (to say 0.99 s ago), and the computation will be repeated.

The above inference problem does not have a unique solution, because there are many sequences of muscle activations that could have caused the state transition we are trying to explain. Even at the final time, the arm could be in many postures that all correspond to a successful grasp; thus the fictive measurement is incomplete. The same ill-posedness is present in the control problem and is known as motor redundancy (Bernstein, 1967). Inference problems do not normally involve this kind of redundancy. Indeed, the inference here is rather unusual: There is a period of time (1 s in our example) when there are no sensory measurements, and the only available measurement at the end of the movement is incomplete. We could consider a different control problem that corresponds to a more usual inference problem involving complete sensory measurements. That control problem is one in which we are given a detailed goal state at each point in time, that is, a reference trajectory for all musculoskeletal degrees of freedom, and have to generate muscle activations so as to force the plant to track this trajectory. When the latter control problem is mapped into an inference problem, the sequence of detailed goal states turns into a sequence of complete sensory measurements, thus eliminating redundancy. It is important to realize, however, that trajectory tracking represents only a small fraction of ecologically relevant behaviors (Todorov & Jordan, 2002). Thus the natural control problem (which involves a large amount of redundancy) corresponds to an unnatural inference problem (in which sensory data are very sparse) and vice versa. Inference is easier if complete sensory measurements are available at all times; similarly, control is easier if detailed goal states are specified at all times.

This reasoning suggests that control is a harder problem than inference, at least in the temporal domain. Indeed, inference in the absence of measurements is called prediction (except that here it is performed backward in time), and prediction tends to be hard. On the other hand, redundancy makes it possible to be sloppy most of the time and still achieve the goal. This is because, even if the initial part of the movement somehow goes wrong, there is time later in

EMANUEL TODOROV Department of Cognitive Science, University of California San Diego, San Diego, California

the movement to observe what happened and take corrective action. The analog of this property in the inference domain is that long-term predictions tend to be inaccurate, while short-term predictions (which correspond to motor commands close to the goal) are more accurate.

The above transformation from control to inference has been instantiated in formal models (Attias, 2003). This is done by setting up a dynamic belief network (see below) that represents the states and actions at different points in time, treating the goal state as being observed, and performing Bayesian inference to find the actions. The shortcoming of this approach is that inference is performed over the product space of states and actions, which for a typical motor control problem is prohibitively large. In the rest of the chapter, we will pursue a different approach, in which the control problem will turn out to be equivalent to an inference problem involving only states. Actions will be defined implicitly as transitions between inferred consecutive states.

Now let us ask the opposite question: Can we start with an inference problem and transform it into a control problem? Consider the problem of estimating the current state of the world, given our previous estimate and the current sensory measurement. One way to do this is to use a predictor-corrector method: Combine the previous estimate with a model of one-step dynamics to obtain a prediction of the current state, and then correct the prediction so as to make it more compatible with the current measurement. The corrected estimate will achieve some tradeoff between being close to the prediction and agreeing with the measurement. The corresponding control formulation is as follows. The entity being controlled (internally) is the state estimate. The control signal corresponds to the correction needed to achieve better agreement with the measurement. Suppose the control is chosen so as to minimize a sum of two costs: an energy cost and an accuracy cost. The energy cost is minimal when there is no correction. The accuracy cost is minimal when the correction is complete. The control that minimizes the sum of these costs will lie somewhere in between the two extremes, thus achieving a similar tradeoff as the predictor-corrector method.

We now see that in the spatial domain, inference can be harder than control. This is because the estimator "controls" (i.e., corrects) all aspects of the estimated state. In the coffee-drinking example, the estimator may deal not only with the arm and the cup, but also with the picture on the wall, the mountains we can see through the window, and many other things that have no relevance to motor actions. Note, however, that this implies a somewhat outdated view of perception in which all aspects of the sensory input are processed in parallel on equal footing. In reality, perception may be geared toward serving the needs of the ongoing behavior, which in our example means ignoring the picture and the mountains and focusing on the arm and cup. In the latter view, inference and control have similar spatial complexity in terms of what needs to be computed (state estimate versus control signal). However, the input to this computation (sensory data versus task goal) is always higher-dimensional for the sensory system.

The above transformation from estimation to control corresponds to the idea of minimum-energy filtering (Mortensen, 1968), in which estimation is formulated as a minimum-energy tracking problem and is solved by using optimal control methods. The shortcoming of this approach is that it yields only point estimates, while a lot of evidence (see below) indicates that the brain computes probability distributions rather than point estimates. In the rest of the chapter, we will pursue a different approach in which the estimator computes the full Bayesian posterior.

To summarize the main points in this section, the similarities between control and estimation arise when task goals are associated with sensory measurements and control signals are associated with corrections to the state estimate. Our discussion was framed in the context of optimal control and optimal/Bayesian inference, which was not a coincidence. Indeed, we will see below that optimality is the source of these similarities.

Duality of Bayesian inference and optimal control in isometric tasks

Here, we provide a concrete example illustrating the duality between Bayesian inference and optimal control. Let $\mathbf{u}$ be a vector of muscle activations, $\mathbf{t}$ the resulting vector of joint torques, and M the matrix of moment arms that maps muscle forces (proportional to muscle activations under isometric conditions) to joint torques: $\mathbf{t} = M\mathbf{u}$. *Isometric* means that there is no movement. Since there are more muscles than joints, a desired torque $\mathbf{t}^*$ can be achieved by infinitely many muscle activations $\mathbf{u}$. This is a manifestation of motor redundancy. To select one out of all possible $\mathbf{u}$, the motor system needs some selection criterion. Suppose that this criterion is to keep the sum of squared muscle activations as small as possible. Then $\mathbf{u}$ can be found by minimizing the cost function:

$$\tfrac{1}{2}\left\|\mathbf{t}^* - M\mathbf{u}\right\|^2 + \tfrac{1}{2}r\left\|\mathbf{u}\right\|^2 \qquad (1)$$

The first term is an accuracy cost that is minimal when the desired torque is exactly achieved. The second term is an energy cost that is minimal when all muscle activations are zero. The parameter r determines the relative importance of the two. Quadratic cost functions are usually scaled by a factor of ½ for convenience. We have shown (Todorov, 2002) that the above cost function as well as more realistic versions of it predict the empirical phenomenon of cosine tuning, that is, the fact that muscle activation varies with the

cosine of the angle between the mechanical pulling direction of the muscle and the direction of end-effector force (Hoffman & Strick, 1999).

We now turn to the corresponding inference problem, which involves a Gaussian prior over the elements of $\mathbf{u}$ with mean 0 and variance $1/r$ and a fictitious measurement corresponding to goal achievement, namely, $\mathbf{y} = \mathbf{t}^*$. Bayesian inference requires a generative model, that is, a model of how the (noisy) sensory measurement was generated given the state. In this case, the generative model is $\mathbf{y} = M\mathbf{u} + \boldsymbol{\varepsilon}$, where the elements of $\boldsymbol{\varepsilon}$ are Gaussian with mean 0 and variance 1. Applying Bayes rule (see below) and using the formula for a Gaussian, we obtain the posterior probability of $\mathbf{u}$ given $\mathbf{y}$:

$$p(\mathbf{u}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{u})\, p(\mathbf{u}) \propto \exp\left(-\tfrac{1}{2}\left\|\mathbf{t}^* - M\mathbf{u}\right\|^2 - \tfrac{1}{2}r\|\mathbf{u}\|^2\right) \quad (2)$$

Thus the posterior probability in the inference problem (equation 2) coincides with the exponent of the negative cost in the control problem (equation 1); in particular, the most probable muscle activations coincide with the optimal muscle activations.

This completes our example of duality in isometric tasks. Although it is a simple example that does not involve state variables changing over time, it nevertheless illustrates a key idea that is used extensively later. The idea is that costs and probabilities are related by an exponential transformation. This is to be expected; costs add while probabilities multiply, and it is the exponential transformation that turns sums into products. The same transformation shows up in other fields as well. In statistical mechanics, for example, the energy of a given state and the probability of finding the system in that state at thermal equilibrium are related by the Gibbs distribution, which is the exponent of the negative energy.

We are now ready to develop a general form of duality between optimal control and optimal/Bayesian inference over time. To this end, we will first review the concepts of optimality in sensory and motor processing and note the similarities and differences between the two formalisms. This analysis will then indicate how the control problem should be phrased so as to become mathematically equivalent to Bayesian inference.

Optimality in sensory and motor processing

While all aspects of neural function have evolved to produce behavior that is beneficial to the organism, the evolutionary pressures on real-time sensory and motor processing may have been particularly strong and direct because of the crucial role that such processing plays in getting food to the mouth, escaping predators, and generally keeping the organism alive. It is, then, not surprising that the underlying neural mechanisms perform about as well as any information-processing system that is subject to the same constraints could perform—in other words, near optimally. Indeed, optimality is becoming the theoretical framework of choice in studying both sensory and motor systems (Todorov, 2004; Kording & Wolpert, 2006; Doya, Ishii, Pouget, & Rao, 2007).

In the sensory domain, optimality corresponds to Bayesian inference. In the simplest setting, it involves three probability distributions over the (relevant) state of world: the prior, the likelihood, and the posterior. They are related according to Bayes' rule:

$$\text{posterior}(x) \propto \text{likelihood}(y|x)\,\text{prior}(x) \quad (3)$$

The prior summarizes everything that we know about the state of the world before observing the measurement. The likelihood (which formalizes the generative model) is the probability of measurement y being generated when the world is in state x. The posterior summarizes everything that we know after the measurement is taken into account. If there are multiple independent measurements, the right-hand side of equation 3 contains the product of the corresponding likelihoods. The latter setting is used in models of cue integration, in which subjects are presented with two (often incompatible) sensory cues and asked to estimate some property of the world. Such experiments have provided the simplest and perhaps most compelling evidence that perception relies on Bayesian inference (e.g., Ernst & Banks, 2002). The probability distributions that are used in these studies are typically Gaussian.

Unlike the static nature of many cue integration experiments, sensory processing in the real world takes place in time and requires integration of measurements obtained at different points in time. This is called recursive estimation or filtering. The basic update scheme that is applied at each point in time has the predictor-corrector form:

$$p(x) \propto l(y|x) \sum_{x_{prev}} d(x|x_{prev})\, p(x_{prev}) \quad (4)$$

Here, $p(x)$ is the posterior at the current state, $p(x_{prev})$ is the posterior at the previous state (which we have already computed at the previous time step), $l(y|x)$ is the likelihood function, and $d(x|x_{prev})$ is the stochastic one-step dynamics of the world. In estimating the state of the body, the dynamics d will also depend on the control signal that is available to the sensory system in the form of an efference copy. The product of d and p, which is being summed over, is the joint probability of x and x_{prev}. The sum marginalizes out x_{prev} and yields a prediction (or prior) over x. In this way, the posterior at one point in time is used to compute the prior at the next point in time. The multiplication by the likelihood l is the sensory-based correction discussed earlier.

A number of experimental findings support the notion of Bayesian inference over time (Wolpert, Gharahmani, & Jordan, 1995; Kording & Wolpert, 2004; Saunders & Knill,

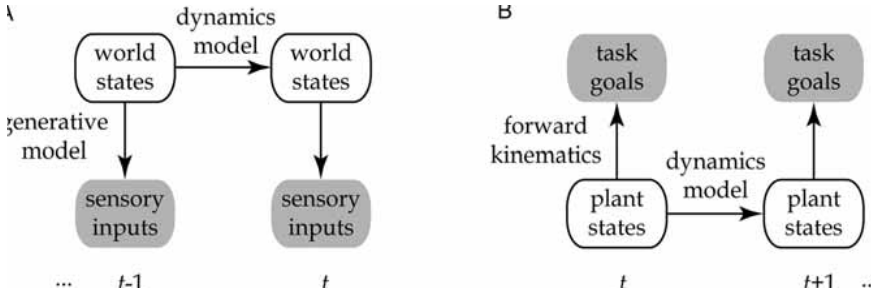


FIGURE 42.1 Belief networks for Bayesian inference and optimal control. (*A*) Shaded nodes correspond to observed quantities; open nodes correspond to random variables whose (marginal) probabilities are to be computed. The dynamics model is the probability distribution of the next state given the current state. The generative model is the probability distribution of the sensory input given the state. (*B*) The optimal control problems in this chapter are mathematically equivalent to Bayesian inference problems; thus they can be represented with belief networks. The forward kinematics play the role of a generative model and indicate whether the goal is achieved by the current state of the plant. The actual generative model specifies a probability distribution proportional to $\exp(-q(x))$, where $q(x)$ is the state cost.

2004). These studies typically use arm movements, not so much for the purpose of studying the motor system but as a continuous readout of perception. Such studies demonstrate that subjects take into account multiple sources of information over time (visual and proprioceptive, along with internal predictions) and rely on that information to guide movements. As in cue integration, the probability distributions that are assumed here are typically Gaussian. When the dynamics are linear and all noise is Gaussian, the posterior is also Gaussian and can be computed by using the Kalman filter.

There is a graphical representation of Bayesian inference problems (figure 42.1*A*) that is known as a graphical model or a belief network (dynamic belief network when time is involved). This representation is very popular in statistics and machine learning (Pearl, 1988). Belief networks help us to understand the mathematical models intuitively and will also be useful later in clarifying the relationship between estimation and control. To avoid confusion, keep in mind that unlike neural networks, the nodes in belief networks do not correspond to neurons, and the arrows do not correspond to synaptic connections. Instead, the nodes correspond to collections of random variables, whose probabilities are presumably represented by populations of neurons. Strictly speaking, the arrows encode conditional probabilities, but in reality they often correspond to the causal relations in the world, as illustrated in figure 42.1*A*. We show only part of the network containing the states of the world at two consecutive points in time as well as the corresponding sensory measurements/inputs. Solid gray circles denote variables whose values are observed and that therefore contribute a likelihood function. Open circles denote variables whose values are to be inferred. The forward arrows encode the stochastic dynamics of the world, that is, the one-step transition probability *d*. The downward arrows encode how sensory measurements are generated as a function of world states. This generative model may incorporate a model of optics in vision or a model of acoustics in audition plus a model of sensory transduction in the corresponding modality. One can think of perception as a computational process that inverts the generative model in a probabilistic sense (this idea goes back to Helmholtz).

Optimality has also been applied in motor control, perhaps even more extensively than in perception. This may be because, apart from its general appeal as an organizing principle, optimality appears to be the right way to resolve redundancy. There is a wealth of experimental data (for reviews, see Todorov, 2004; Kording & Wolpert, 2006) suggesting that the motor system generates actions that maximize task performance or utility. Optimal control models have accounted in parsimonious ways for numerous features of motor behavior on the levels of kinematics, dynamics, and muscle activity. There are two general approaches: open loop control and closed loop control. Open loop control precomputes the entire sequence of motor commands from now until the goal is achieved, while closed loop control (or feedback control) computes only the current motor command given the current state estimate and then uses information about the next state to compute the next command. Since movements are under continuous sensory guidance, the latter type of model corresponds more closely to what the brain does. Although optimal feedback controllers are harder to construct, we now have efficient algorithms and fast computers that enable us to explore such models.

Here is how optimal feedback control works in a nutshell: Define an instantaneous cost that accumulates over time and yields a cumulative cost. The instantaneous cost is usually a sum of a control cost $r(u)$, which encourages energetic efficiency, and a state cost $q(x)$, which encourages accuracy or, more generally, getting to desirable states and avoiding

undesirable states. Also define the one-step stochastic dynamics $d_u(x_{next}|x)$, which is similar to the one-step transition probability in Bayesian inference except that it now depends on the control u explicitly. The objective is to find an optimal control law, that is, a mapping from states to controls that minimizes the expected cumulative cost. This computation is facilitated by the optimal cost-to-go function $v(x)$, defined as the cost that is expected to accumulate if the plant is initialized at state x and is controlled optimally thereafter. The optimal cost-to-go function plays a key role because it summarizes all relevant information about the future and allows us to compute the optimal control at the present time using greedy optimization without look-ahead. This function is the unique solution to the Bellman equation:

$$v(x) = \min_u \left[q(x) + r(u) + \sum_{x_{next}} d_u(x_{next}|x) v(x_{next}) \right] \quad (5)$$

The control u that achieves the minimum is the optimal control at the current state x. This equation is quite intuitive; it says that the optimal cost-to-go can be broken down into the instantaneous cost incurred at the current state when applying the optimal control plus the optimal cost-to-go for the movement originating at the next state. The expectation (sum over x_{next}) is needed because the dynamics are stochastic and the next state is known only in a probabilistic sense. Equation 5 can always be solved numerically by using dynamic programming, which involves computing the minimum over u and assigning it to $v(x)$ for each x and each time step. For large problems, however, this computation is often intractable.

One special case in which equation 5 can be solved efficiently is the case of linear dynamics, Gaussian noise, and quadratic costs. In such problems (known is LQG), the optimal cost-to-go is quadratic and can be computed with a method very similar to the Kalman filter. Note that a Gaussian is the exponent of a quadratic function, the product of two Gaussians is the exponent of the sum of the corresponding quadratics, and a sum of quadratics is again a quadratic (as illustrated in the isometric task example). So both the Kalman filter and the LQG optimal controller are based on manipulating quadratics; indeed, the underlying equations are identical. This duality was discovered by Kalman (1960) and was the first indication that optimal estimation and optimal control are closely related. We recently showed (Todorov, 2008) that Kalman's duality is special to the LQG setting and does not generalize. However, there exists another form of duality that does generalize. It was developed by Mitter and Newton (2003) in continuous time and by Todorov (2006, 2008) in discrete time. The two developments are technically quite different yet yield related results. Our presentation in the next section will use the discrete-time version, which is more intuitive and also turns out to be more general. The continuous-time version can be obtained as a special case by assuming Gaussian noise and taking a certain limit.

General duality of Bayesian inference and optimal control

Comparing equations 4 and 5, we can already see a similarity between optimal control and Bayesian inference. The state cost q and the likelihood l are related in the sense that they both inject new information about x in each step of the recursive process. The optimal cost-to-go v and the posterior p are related in the sense that they both accumulate information about x over time. The one-step dynamics d are present in both control and estimation. One equation involves sums, while the other involves products, but sums can be turned into products by the exponential transformation. Yet we also see a difference: While equation 4 specifies the posterior p directly via an explicit formula, equation 5 specifies the optimal cost-to-go v indirectly as the solution to an unsolved optimization problem. Neither the control cost $r(u)$ nor the dependence of d_u on u has an analog in 4, suggesting that whether or not the optimal control problem is dual to a Bayesian inference problem will depend on how we define $r(u)$ and d_u.

To establish a general duality, we will define the control signal as a probability distribution over possible next states. This is unusual but in retrospect natural. What controls do is affect the plant dynamics. So we can characterize them directly in terms of how they affect the plant dynamics. For a stochastic plant, this characterization takes the form of a probability distribution $u(x_{next})$. Thus the one-step dynamics are simply

$$d_u(x_{next}|x) = u(x_{next}) \quad (6)$$

According to this definition, the controller has the power to impose on the plant whatever dynamics it wishes. We will restrict this power somewhat, by defining the passive dynamics $d(x_{next}|x)$ in the same way as in the inference problem and allowing u to be nonzero only if d is nonzero. The passive dynamics capture the effects of gravity, interaction forces, and motor noise. The above restriction means that the control signals can cause only those state transitions that could have occurred by accident, that is, the noise and the controls are restricted to act in the same subspace. For musculoskeletal plants in which the controls correspond to muscle activations, the noise model should be restricted to muscle space and should not be allowed to act directly on, say, arm position (which would be physically unrealistic anyway). The above restriction also means that we cannot model external perturbations acting on objects of interest. Such perturbations are often used experimentally to probe

the visual feedback control laws; however, they are uncommon in the real world.

Having a model of passive or uncontrolled dynamics allows us to define the control cost in a natural way. Intuitively, such a cost should measure how large the control signals are. Larger control signals have larger effects on the plant dynamics; that is, they push the plant farther away from its passive dynamics. This suggests a control cost that measures the difference between the probability distributions $u(x_{next})$ and $d(x_{next}|x)$. Differences between probability distributions are most commonly measured by using Kullback-Liebler (KL) divergence; thus the control cost will be defined as

$$r(u) = \mathrm{KL}(u,d) = \sum_{x_{next}} u(x_{next}) \log \frac{u(x_{next})}{d(x_{next}|x)} \tag{7}$$

Definitions 6 and 7 yield a family of control problems that still satisfy the Bellman equation 5 but have additional structure that can be exploited. A control problem in our family is defined by specifying the state cost $q(x)$ and passive dynamics $d(x_{next}|x)$. Once q and d are given, we can substitute equations 6 and 7 into equation 5 and observe that the minimization with respect to u can be performed analytically, owing to properties of the KL divergence. We omit the derivation and summarize the results. The results are expressed most conveniently in terms of a desirability function defined as

$$z(x) = \exp(-v(x)) \tag{8}$$

When the optimal cost-to-go $v(x)$ is small, the function $z(x)$ is large, hence the term *desirability*. It also rhymes with *probability*, which is appropriate because z will turn out to behave like a probability distribution. It can now be shown that the optimal control (i.e., the optimal next-state probability distribution) at state x is

$$u^*(x_{next}) \propto d(x_{next}|x)\, z(x_{next}) \tag{9}$$

This form of control is illustrated in figure 42.2. Given the current state x, we multiply the one-step passive dynamics $d(x_{next}|x)$ by the desirability $z(x_{next})$, normalize to obtain a proper probability distribution $u^*(x_{next})$, and sample the next state from it. Note that multiplication by the desirability z has the effect of shifting the passive dynamics d toward more desirable states.

We still need to compute z. This is done by substituting the optimal control (equation 9) into the Bellman equation (equation 5), dropping the min operator, and exponentiating so as to obtain an update for z rather than v. The resulting update is

$$z(x) = \exp(-q(x)) \sum_{x_{next}} d(x_{next}|x)\, z(x_{next}) \tag{10}$$

The similarity with Bayesian inference (equation 4) is now obvious: The desirability z corresponds to the posterior probability p, the exponentiated state cost $\exp(-q)$ corresponds to the likelihood l, and the one-step transition probability d plays the same role in both cases. The only difference is that z is updated backward in time, while p is updated forward in time. This is because control is about the future, while inference is normally about the past. However, if we construct the inference problem as outlined earlier, that is, provide fictive sensory measurements in the future, then Bayesian inference and optimal control become mathematically equivalent.

Optimal control can then be represented with the belief network shown in figure 42.1*B*. This network is drawn upside-down so as to highlight an important difference between inference and control. In inference, the known quantities (sensory measurements) are near the periphery, while in control, the known quantities (task goals) are deep inside the CNS. Conversely, the outputs of the sensory system are deep inside the CNS, while the outputs of the motor system are close to the periphery. Both diagrams are oriented so that the CNS is up and the periphery is down. One network involves "world states," and the other involves "plant states"; however, these two notions of state may actually be similar. This is because the motor system has to represent not only the state of the body (plant), but also all relevant aspects of the state of the environment, while the sensory system may not represent all aspects of the world but instead may focus on those relevant to the ongoing behavior.

Thus far, estimation and control have been discussed separately, while in the brain they are performed simultaneously. Can we think of both sensory and motor processing as being part of the same computation? This can be done by combining the two belief networks in figure 42.1 and

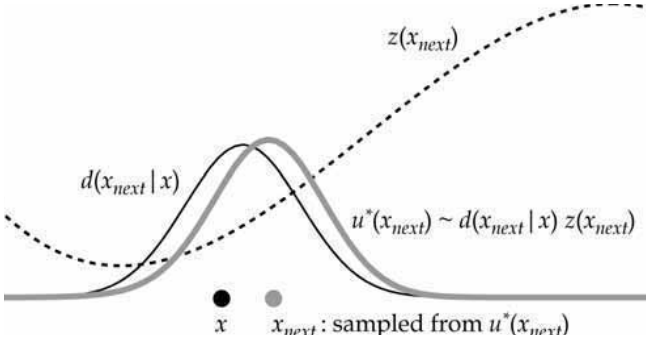


FIGURE 42.2 Optimal control with probability distributions. The passive dynamics $d(x_{next}|x)$ is the probability distribution of the next state when the system is not controlled. The control $u(x_{next})$ is the probability distribution of the next state when the system is controlled. The optimal control $u^*(x_{next})$ is proportional to the product of the passive dynamics $d(x_{next}|x)$ and the desirability function $z(x_{next})$. Multiplying a narrow probability distribution by a smooth function has the effect of shifting the distribution along the gradient of that function. Thus the optimal control is similar in shape to the passive dynamics but is shifted towards more desirable states.

performing Bayesian inference on the composite network. The sensory measurements in the past would be real, while those in the future would be fictive. The probability over past states would encode what we believe has already happened, while the probability over future states would encode what we believe will happen if we act optimally. The control problem was set up in such a way that having a prediction about the future is equivalent to specifying a control signal that turns this prediction into reality. Note that such a unified computational scheme would be only approximately optimal, because the controller here was designed with the assumption that the current state is known with certainty. This form of approximation tends to be quite accurate and is often used in control engineering (it is known as certainty equivalence control). The approximation fails when the uncertainty about the state affects the optimal actions, as in tasks that involve tradeoffs between exploration and goal achievement. In that case, the state in the control problem can be augmented with the uncertainty about the state in the inference problem (Simpkins, de Callafon, & Todorov, 2008).

To summarize this section, we described a family of optimal control problems that are mathematically equivalent (i.e., dual) to Bayesian inference. The state costs and the passive dynamics in our formulation are completely general and can be defined in whatever way is necessary. The only constraints are that the control signals must act in the same subspace as the passive dynamics and the control cost must equal the KL divergence between the controlled and passive dynamics. In the continuous-time limit, this control cost reduces to the familiar quadratic energy cost. Control problems that do not satisfy the above constraints do not seem to have exact duals, yet they can often be approximated with problems that satisfy the constraints (Todorov, 2006).

Intermediate representations: Sensory features and motor synergies

On the system level, sensory processing performs a transformation from sensory inputs to inferred states, while motor processing performs a transformation from task goals to motor commands. However, neither transformation is performed monolithically by a single brain area. Instead, multiple brain areas are involved, and most of them use neural representations that correspond to neither the input nor the output of the overall computation but to something in between. Similarly, if we analyze a typical computer program, we will notice that most of the variables that are declared in it are internal variables that represent intermediate results. How do such intermediate representations relate to the overall computation? One way to address (or rather avoid) this question is the computer science way, adapted to neuroscience by Marr (1982). In this approach, one makes a strict distinction between the problem being solved, the algorithm for solving it, and the implementation of the algorithm in software, hardware, or wetware. While this approach has many merits, a significant drawback is that computational-level analyses tell us little about the underlying neural representations and the interactions among them. In this section, we outline a somewhat different approach that enables us to relate intermediate representations to the overall computation more directly. We will first develop the idea for sensory systems and then see how it applies to motor systems.

Intermediate sensory representations are often called "features" and are thought to be features of the "stimulus." But what is a stimulus? Is it the sensory input, or is it the relevant aspect of the world reflected in the input? If features are defined as functions of the sensory input, then they do not belong on the computational level, and we are back to Marr's strict separation. Suppose instead that features are statements about the state of the world. For example, suppose that the activity of an "edge detector" in primary visual cortex is not a statement about the presence of an edge in the retinal image, but a statement about the state of the world that caused the retinal image to contain an edge. In this view, features are part of the generative model (figure 42.3*A*). The sensory input is modeled as a (probabilistic) function of the features instead of the other way around. Bayesian inference can be applied to such a hierarchical generative model without modification. One prediction is that at every intermediate level of sensory processing, there will be both bottom-up and top-down effects. This is because the probability of any variable in a belief network generally depends on all other variables. The beauty of this approach is that different levels of the generative model can be instantiated in different brain areas, and as long as the communication within and between areas corresponds to Bayesian inference, the entire distributed system will perform a single computation, using perhaps a single algorithm (see below) that operates in parallel on multiple representations.

Let us now apply the same idea to the motor system. The closest analog of a sensory feature in motor control is the notion of a motor synergy. It corresponds to some intermediate representation that is more abstract than the full musculoskeletal state but more detailed than the task goal. By analogy to the sensory system, we propose that synergies are part of a hierarchical generative model, which in the case of the motor system is a (probabilistic) mapping from plant states to task goals. Synergies are often thought to be related to motor commands rather than plant states; however, recall that in our formulation, motor commands are implicit and can be recovered from the probability of the future states under the optimal controls. As illustrated in figure 42.3*B*, synergies can be used for both spatial and temporal abstraction. For example, a synergy might be a statement about the shape of the fingertip trajectory over a short period of time.

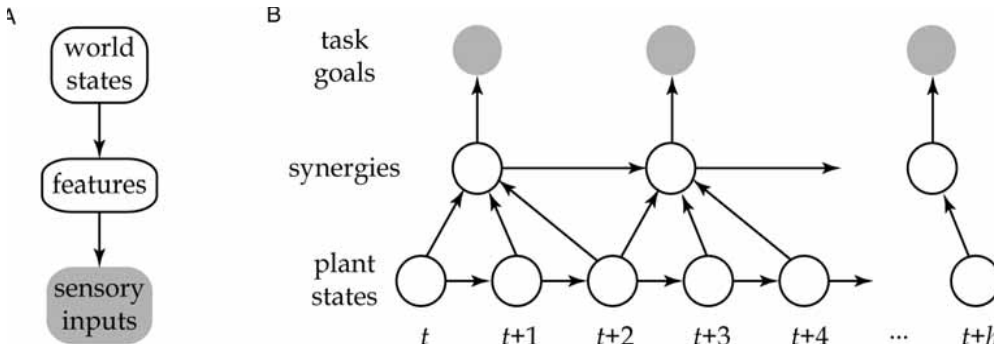


FIGURE 42.3 Belief networks for hierarchical Bayesian inference and optimal control. (*A*) By defining intermediate sensory representations (features), we can construct hierarchical generative models. The features become part of the model of how sensory inputs depend on states of the world. They are not needed to build generative models but presumably facilitate the inversion of such models using Bayesian inference. (*B*) By defining intermediate motor representations (synergies) that depend on only some aspects of the state but over extended periods of time, we can achieve both spatial and temporal abstraction. This is done by using cost functions of the form $q(h(x_t \ldots x_{t+d}))$, where h are the synergy states and d is the temporal abstraction horizon. The synergies become part of the model of how goal achievement depends on the plant state. Control is about achieving goals that are removed in time, which requires unfolding the time axis and representing multiple time steps. Limiting this unfolding to a fixed number of steps into the future is called receding horizon control. At the horizon $t + h$, we need some approximation of the desirability function $z(x_{t+h})$.

The fact that the synergy corresponds to a period of time and not a single point in time yields temporal abstraction. The fact that the synergy corresponds to only some aspects of the state of the plant and not the entire state (e.g., it does not specify all the joint angles but only the fingertip position) yields spatial abstraction. Different forms of spatial and temporal abstraction have played an important role in designing automatic controllers for complex tasks, suggesting that the brain may also rely on such tools.

Thus intermediate representations in both sensory and motor systems can be thought of as being part of hierarchical generative models. One might ask, however, what is the point of having such representations when generative models can be built without them. For example, given the full state of the arm, we can directly compute where the fingertips are, without the help of motor synergies. Similarly, we can directly compute the retinal image resulting from a given configuration of three-dimensional objects and light sources, without relying on sensory features (this is what computer graphics does). Indeed, intermediate representations not only are unnecessary to build generative models, but may even complicate the construction of such models. However, the goal of both the sensory and motor systems is not so much to build generative models but rather to invert them. The inversion is the harder problem and is also the problem that has to be solved in real time. Intermediate representations are likely to facilitate this inversion, by providing various forms of abstraction and enabling the inference algorithm to construct the final answer in manageable pieces. Thus intermediate representations may exist not for the sake of representation but because they facilitate the computation.

One might also ask where intermediate representations come from. In sensory systems, it has been shown that unsupervised learning applied to collections of natural sensory inputs can recover the features that are observed experimentally. The most notable examples come from the visual system (Olshausen & Field, 1996), although the approach has also been applied successfully to the auditory system (Lewicki, 2002). Unsupervised learning looks for statistical regularities in high-dimensional data. Traditional unsupervised learning methods such as principal components analysis reduce the dimensionality of the data. In contrast, the forms of unsupervised learning that are thought to be used by sensory systems tend to increase dimensionality, that is, they form overcomplete (and sparse) representations. This might seem counterproductive; however, it resonates well with recent computational approaches in which increasing dimensionality simplifies computation. Support vector machines and kernel methods in general are based on this idea (Scholkopf & Smola, 2001). Liquid state machines in neuroscience have the same flavor (Maass, Natschlager, & Markram, 2002).

Unsupervised learning has also been applied in motor control to extract candidate synergies (D'Avella, Salticl, & Bizzi, 2003; Santello, Flanders, & Soechting, 1998). However, the situation here is qualitatively different. While in sensory systems, unsupervised learning is applied to sensory data that are available to the brain during learning/development, in motor systems, it is applied to movement data that are available to the brain only after it has mastered the motor task. If we agree that appropriate synergies must exist before successful movements can be generated in a given task, then the brain cannot learn those synergies from successful move-

ments. In other words, the unsupervised learning methods that are used by motor control researchers are not a feasible model of learning by the motor system. A feasible model should learn on the basis of information that is available at the time of learning. The one thing that is always available is the input, which in the case of the motor system corresponds to the task goals. Thus the analog of learning features from sensory inputs would be learning synergies from task goals. Unfortunately, the task goals are not directly accessible to an external observer, so the application of unsupervised learning as outlined here is not easy; yet we suspect that it is worth pursuing.

Another insight into motor synergies that comes from the analogy with sensory systems is the number of synergies. It is widely believed that motor synergies serve the purpose of dimensionality reduction; indeed, they are usually defined as the outputs of dimensionality reduction algorithms. However, as was discussed earlier, dimensionality expansion rather than reduction may be more beneficial in terms of simplifying computation. Furthermore, the number of different neural activation patterns in, say, primary motor cortex greatly exceeds the number of musculoskeletal degrees of freedom. If we agree to think of neural activity in motor areas as representing synergies, in the same way that we think of neural activity in sensory areas as representing features, the dimensionality expansion point of view becomes unavoidable.

This view represents a significant departure from the established thinking about motor synergies and might at first seem incompatible with the evidence that large amounts of variance (in movement kinematics, electromyelograms, or isometric forces) can be explained by small numbers of components. How can behavioral evidence for dimensionality reduction be reconciled with intermediate representations performing dimensionality expansion? One answer comes from our work on optimal control (Todorov & Jordan, 2002; Todorov, 2004), in which we showed that an optimally controlled redundant system will exhibit signs of dimensionality reduction regardless of how the controller is implemented. If that is the case and the motor system is good at approximating optimal controllers, then a lot of the dimensionality reduction results that are currently taken as evidence for synergies might instead be indirect evidence for optimality. The overcomplete intermediate representations that we propose to call synergies may be the mechanism that enables the motor system to perform near optimally.

Algorithms for learning and online computation

Bayesian inference and optimal control are of interest in many fields (e.g., statistics, computer science, signal processing, control engineering, economics). Consequently, many algorithms have been developed. While none of them can yet compete with biological sensory and motor systems on complex real-world problems, this repository of algorithmic knowledge is an important source of insights into what the brain might be doing. We refer the reader to the work of Doya and colleagues (2007) for an extended discussion. Here, we make only a few points that are relevant to this chapter.

One class of Bayesian inference algorithms, known as belief propagation (Pearl, 1988), are reminiscent of computation in recurrent neural networks except that the messages that are being exchanged are probability distributions (presumably encoded by populations of neurons). The analog in optimal control is dynamic programming, which for the family of control problems described above is reduced to belief propagation. An important corollary of the estimation control duality is that sensory population codes that are thought to represent probability distributions (Doya et al., 2007) can be equally useful in motor control. Both belief propagation and dynamic programming are global methods in the sense that they aim to compute functions over the entire state space. For large problems, this is unlikely to be doable in real time. In the control domain, this point is well appreciated; indeed, dynamic programming is normally applied off-line so as to precompute/learn the optimal control law. The latter is then used online to generate motor commands as a function of plant states and goal parameters (e.g., target positions). The equivalent in Bayesian inference would be to learn a direct mapping from sensory inputs to state estimates, which is not how people usually think about inference. There may be several reasons for this: (1) The input to the sensory system is so high-dimensional that learning such a mapping is infeasible; (2) inference is an easier problem than control (recall our discussion of redundancy), so the computation is easier to perform online; and (3) the brain actually learns direct mappings from sensory inputs to estimated states, but this is not yet reflected in most Bayesian inference algorithms. One exception here is the Helmholtz machine (Dayan, Hinton, Neal, & Zemel, 1995), which is a belief network augmented with a mechanism for learning the direct mapping discussed above (i.e., the inverse of the generative model).

Regardless of whether and how much of the transformation is learned, it is clear that a lot of processing takes place in real time in both the sensory and motor systems. There is a simple way to combine the advantages of learning and on-line computation: Learn a global but approximate transformation, use it on-line for initialization, and then apply an on-line algorithm to refine the solution locally around the current state. Locally, probabilities and costs can be approximated by simplified models (e.g., Gaussians and quadratics), which afford faster computation. Such local approximation

methods are available in both estimation (the extended Kalman filter) and control (iterative LQG or differential dynamic programming). In the case of control, local improvement requires unfolding the time axis up to a certain horizon. This is known as receding horizon control. For our family of control problems, it is illustrated with the belief network in figure 42.3*B*. If the motor system relies on such methods, we should expect to find neurons coding the state of the plant at multiple points in time in the future. Indeed, it has often been noted (e.g., Kalaska, Sergio, & Cisek, 1998) that the latency between neural firing and motor behavior has a broad distribution and, on average, is substantially longer than what one would expect from conduction latencies alone. Such data can tell us how much unfolding is taking place in the motor system. The answer is on the order of 200 ms for reaching movements, although more complex tasks (for which we do not have data) may require unfolding over longer time horizons.

Bayesian inference problems can also be solved by using sampling methods. An example is Gibbs sampling, which works by choosing a node to be updated, resampling its value from the conditional probability given the current values of its neighbors, choosing another node, and so on. After a "burn-in" period, the samples that are generated in this way match the correct Bayesian posterior. The estimation-control duality makes it possible to apply sampling algorithms to optimal control problems as well. Sampling algorithms have not been seriously considered as models of brain function, but perhaps they should be, for several reasons. First, these are the only algorithms that are actually guaranteed to solve the problem (even though it may take a long time). All other algorithms when applied to continuous state variables require function approximation and, as a result, may never converge or may converge to the wrong answer. Second, sampling is inherently parallel. Other algorithms can be parallelized but not to the same extent. This is an important consideration, given the staggering number of neurons in the brain. Third, sampling is inherently stochastic. Implementing it in a deterministic computer requires a pseudo-random number generator. The brain has internal sources of noise (e.g., failures of synaptic transmission) that could be used as random number generators, implying that neural noise may be a feature rather than a nuisance.

In summary, a range of algorithms for Bayesian inference and optimal control have been developed in multiple fields. Furthermore, the estimation control duality makes it possible to take estimation algorithms and apply them to control problems and vice versa. Such algorithms are very relevant to neuroscience because they solve the same problems that the sensory and motor systems appear to be solving. Which of these algorithms resemble the ones used by the brain is not yet clear (but see Doya et al., 2007). Algorithmic issues have generally received limited attention in neuroscience, perhaps because they are hard to address experimentally. This is in contrast to system-level computations, which can be addressed by using behavioral data, and neural representations, which can be addressed by using single neuron data. Indeed, a lot is already known about both system-level computations and neural representations, in both the sensory and motor systems. This knowledge imposes strong constraints, which, in conjunction with algorithmic insights from multiple fields, may soon enable us to go after the brain's algorithms in a systematic way.

ACKNOWLEDGMENTS This work was supported by the U.S. National Science Foundation.

REFERENCES

Attias, H. (2003). Planning by probabilistic inference. In *Proceedings of the 9th International Conference on Artificial Intelligence and Statistics*, held in Key West, FL.

Bernstein, N. (1967). *The coordination and regulation of movements.* Oxford, UK: Pergamon.

D'Avella, A., Saltiel, P., & Bizzi, E. (2003). Combinations of muscle synergies in the construction of a natural motor behavior. *Nat. Neurosci.*, *6*, 300–308.

Dayan, P., Hinton, G., Neal, R., & Zemel, R. (1995). The Helmholtz machine. *Neural Comput.*, 7, 1022–1037.

Doya, K., Ishii, S., Pouget, A., & Rao, R. (2007). *Bayesian brain: Probabilistic approaches to neural coding.* Cambridge, MA: MIT Press.

Ernst, M., & Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Hoffman, D., & Strick, P. (1999). Step-tracking movements of the wrist: IV. Muscle activity associated with movements in different directions. *J. Neurophysiol.*, *81*, 319–333.

Kalaska, J., Sergio, L., & Cisek, P. (1998). Cortical control of whole-arm motor tasks. In M. Glickstein (Ed.), *Sensory guidance of movement* (176–201). Chichester, UK: Wiley.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *ASME Trans. J. Basic Eng.*, *82*, 35–45.

Kording, K., & Wolpert, D. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–247.

Kording, K., & Wolpert, D. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn. Sci.*, *10*, 320–326.

Lewicki, M. (2002). Efficient coding of natural sounds. *Nat. Neurosci.*, *5*, 356–363.

Maass, W., Natschlager, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.*, *14*, 2531–2560.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Mitter, S., & Newton, N. (2003). A variational approach to nonlinear estimation. *SIAM J. Control Optimization*, *42*, 1813–1833.

Mortensen, R. (1968). Maximum-likelihood recursive nonlinear filtering. *J. Optimization Theory Appli.*, *2*, 386–394.

Olshausen, B., & Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco: Morgan Kaufmann.

Santello, M., Flanders, M., & Soechting, J. (1998). Postural hand synergies for tool use. *J. Neurosci.*, *18*, 10105–10115.

Saunders, J., & Knill, D. (2004). Visual feedback control of hand movements. *J. Neurosci.*, *24*, 3223–3234.

Scholkopf, B., & Smola, A. (2001). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, MA: MIT Press.

Simpkins, A., de Callafon, R., & Todorov, E. (2008). Optimal trade-off between exploration and exploitation. In *Proceedings of the American Control Conference* (pp. 33–38), held in Seattle, WA.

Todorov, E. (2002). Cosine tuning minimizes motor errors. *Neural Comput.*, *14*, 1233–1260.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nat. Neurosci.*, *7*, 907–915.

Todorov, E. (2006). Linearly-solvable Markov decision problems. In *Adv. Neural Information Proc. Syst.*, *19*, 1369–1376.

Todorov, E. (2008). General duality between optimal control and estimation. In *Proceedings of the 47th IEEE Conference on Decision and Control* (pp. 4286–4292), held in Cancun, Mexico.

Todorov, E., & Jordan, M. (2002). Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.*, *5*, 1226–1235.

Wolpert, D., Gharahmani, Z., & Jordan, M. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1882.

43 The Mirror Neuron System: A Motor-Based Mechanism for Action and Intention Understanding

GIACOMO RIZZOLATTI, LEONARDO FOGASSI, AND VITTORIO GALLESE

ABSTRACT In this chapter we provide evidence that the cortical motor system is involved in action and intention understanding. In the first part of the chapter, we show that at the core of the cortical motor system, formed by ventral premotor and inferior parietal cortex, there are vocabularies of motor acts, such as grasping, holding, and breaking. Neurons that form these vocabularies code the goal of motor acts independent of how the goal is achieved. Many of these motor neurons also respond to the observation of the same motor acts they motorically code (mirror neurons). In the second part, we show that mirror neurons are involved in both the understanding of motor acts done by others and the understanding of the intention behind the acts. In the last part of the chapter we show that the mirror system in humans also plays a role in action and intention understanding. We conclude by presenting data suggesting that some of the deficits present in the autistic syndrome could be caused by an impairment of the mirror system.

Traditionally, it has been assumed that understanding actions done by others, and even more so their intentions, occurs by applying a kind of reasoning not much different from that used to solve a logical problem. According to this view, when witnessing the actions of others, we process the actions with our sensory system; this information is then elaborated by some sophisticated cognitive apparatus and compared with other similar, previously stored data. At the end of this process, we know what others are doing and why.

Such complex cognitive operation likely occurs in many situations, for example, when the behavior of the observed person is difficult to interpret (Brass, Schmitt, Spengler, & Gergely, 2007). Yet the simplicity and lack of effort with which we usually understand what the others are doing suggest an alternative solution. The actions done by others, after being processed in the observer's visual system, are directly mapped on his or her motor representations without any need of cognitive mediation.

Strong evidence in favor of the existence of a direct mechanism of understanding others' actions by matching them on the observer's own motor system came from the discovery of mirror neurons (MNs), a class of visuomotor neurons that discharge both when a monkey performs goal-related motor acts (e.g., grasping) and when it observes or hears another individual (monkey or human) doing similar acts. Neurons with these properties are found in the rostral sector of the ventral premotor cortex (area F5) and in a sector of the posterior parietal cortex (essentially corresponding to area PFG) that is anatomically connected with area F5. Thus, premotor and parietal MNs form a cortical mirror neuron system that translates sensory information about biological actions into a motor format.

There is evidence that in addition to the parietofrontal mirror neuron system, there are other mirror systems, at least in humans. One, most likely present also in monkeys, is involved in translating observed emotions into a visceromotor pattern that expresses the same emotions (see Gallese, Keysers, & Rizzolatti, 2004). In addition, humans are endowed with a mirror system for phonemes (Fadiga, Craighero, Buccino, & Rizzolatti, 2002) and one for coding non-goal-directed movements (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995).

In the present chapter, we will focus on the parietofrontal mirror system for actions. We will first review the anatomical and functional properties of the mirror system in monkeys and humans and address the issue of how action is represented within primates cortical motor system. We will then discuss a neurophysiological model of how actions and the intentions that promote them are understood. Finally, we will discuss some implications of this model for our understanding of autism.

GIACOMO RIZZOLATTI, and VITTORIO GALLESE Dipartimento di Neuroscienze, Università di Parma, Parma, Italy
LEONARDO FOGASSI Dipartimento di Psicologia and Dipartimento di Neuroscienze, Università di Parma, Parma, Italy

Anatomy of the Mirror Neuron System MNs were first discovered in area F5, which occupies the rostralmost sector of ventral premotor cortex. This region has recently been parcellated (Nelissen, Luppino, Vanduffel, Rizzolatti, & Orban, 2005; Belmalih et al., 2009) into three sectors occupying the cortical convexity (F5c), the posterior bank of the inferior limb of the arcuate sulcus (F5p), and the fundus of the inferior limb of the arcuate sulcus (F5a) (figure 43.1). MNs are generally found in area F5c.

Parietal MNs have been found in the rostral part of the inferior parietal lobule (IPL) convexity (see Rizzolatti & Craighero, 2004), particularly in area PFG (Pandya & Seltzer, 1982; Gregoriou, Borra, Matelli, & Luppino, 2006), and in the anterior intraparietal area (AIP).

Both these areas are connected with the cortex located inside the superior temporal sulcus (STS), including two areas that are selectively active during hand action observation: STPm and LB2 (Nelissen et al., 2005; Perrett et al., 1989). The first one is specifically connected with PFG, while the other, which is also shape sensitive, conveys information to AIP.

Hodological studies (Matelli, Camarda, Glickstein, & Rizzolatti, 1986; Rozzi et al., 2006) showed a reciprocal pattern of connectivity between areas PFG, AIP, and F5. Given the similarity between the functional properties of premotor and parietal MNs, these anatomical data corroborate the idea that areas F5, PFG, and AIP constitute the mirror system for action. As far as the STS areas are concerned, although fundamental for providing visual information on biological motion, they cannot be considered as part of the mirror system in a strict sense, because they do not appear to have motor properties.

Goal-Relatedness and Goal-Chaining in the Ventral Premotor Cortex and in the Inferior Parietal Lobule The functional properties of MNs can be better understood by framing them within the conception that the basic organization of the cortical motor system is in terms of goal-directed movements (Rizzolatti, Luppino, & Matelli, 1998; Rizzolatti, Fogassi, & Gallese, 2000; Crutcher & Alexander, 1990; Alexander & Crutcher, 1990; Kakei, Hoffman, & Strick, 1999, 2001; Hoshi & Tanji, 2000) and not in terms of elementary body part displacements, as was classically thought. Goal-directed movements (i.e., motor acts) are the nuclear building blocks around which action is organized and understood (Rizzolatti et al., 1988; Murata et al., 1997; Raos, Umiltà, Fogassi, & Gallese, 2006; Umiltà et al., 2008).

Particularly important for establishing this concept have been the studies in which single neurons were recorded in a naturalistic context. These studies showed that, typically, the discharge of F5 neurons correlates much better with a motor act than with the movements forming it. Thus many neurons discharge when a motor act (e.g., grasping) is performed with effectors as different as the right hand, the left hand, or the mouth. Furthermore, for the vast majority of neurons, the same type of movement (e.g., an index finger flexion) that is effective in triggering a neuron during a motor act (e.g., grasping) is not effective during another one (e.g., scratching).

By using motor act as classification criterion, F5 neurons were subdivided into various categories such as "grasping-

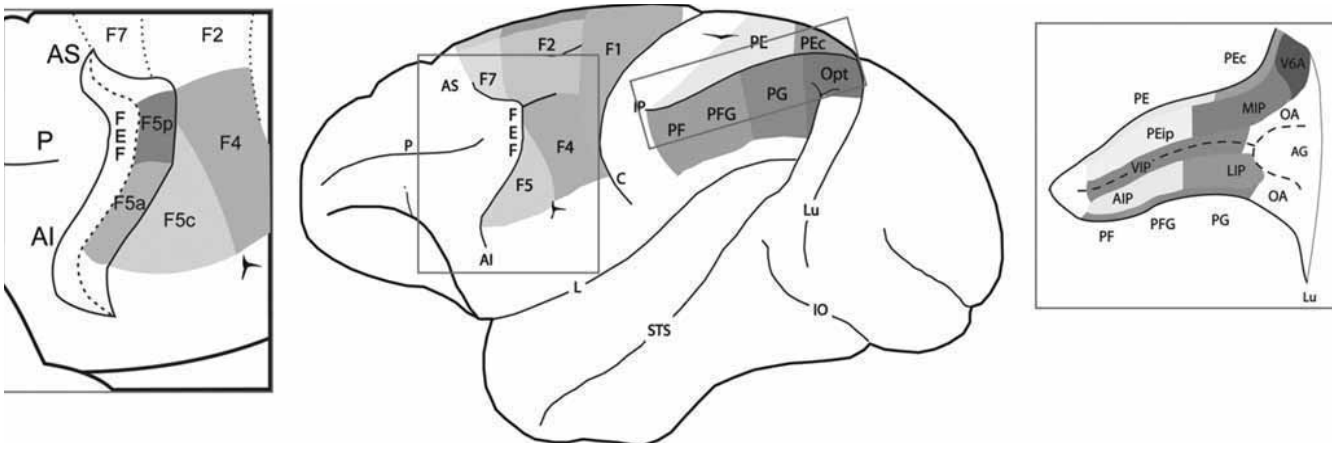


Figure 43.1 Lateral view of the monkey brain showing the parcellation of the motor and the posterior parietal cortex. The areas located within the arcuate and the intraparietal sulcus are shown in an unfolded view of these sulci in the left and right parts of the figure, respectively. For the nomenclature and definition, see Rizzolatti, Luppino, and Matelli (1998), Nelissen et al. (2005), and Gregoriou et al. (2006). Abbreviations: AI, inferior arcuate sulcus; AS, superior arcuate sulcus; C, central sulcus; FEF, frontal eye-fields; IP, intraparietal sulcus; IO, inferior occipital sulcus; L, lateral fissure; Lu, lunate sulcus; P, principal sulcus; STS, superior temporal sulcus. (See color plate 54.)

with-the-hand-and-the-mouth" neurons, "grasping-with-the-hand" neurons, "holding" neurons, "tearing" neurons, and "manipulating" neurons.

In each class, many neurons (about 80%) code specific types of hand shaping, such as precision grip (the grip type most represented), whole hand prehension, and finger prehension. Whether they are specific or not for a certain type of prehension, these neurons show a variety of temporal relations with the prehension phases. Some neurons discharge during the whole motor act, sometimes starting to fire at stimulus presentation. Some other neurons are mostly active during the opening of the fingers, and some are mostly active during finger closure (see Jeannerod, Arbib, Rizzolatti, & Sakata, 1995).

On the basis of these properties, it has been suggested that F5 contains a "vocabulary" (a storage) of *motor act representations*. The vocabulary is constituted by "words," each of which is represented by a set of F5 neurons. Some words indicate the general goal of a motor act (grasping, holding, tearing, etc.). Other words indicate the way in which a specific motor act must be executed (e.g., precision grip or finger prehension). Finally, other words are concerned with the temporal segmentation of the motor act into smaller chunks, each coding a specific phase of the grip (e.g., hand opening, hand closure).

A crucial demonstration of this notion was recently achieved by Umiltà and colleagues (2008). In this study, hand-related neurons were recorded from premotor area F5 and the primary motor cortex (area F1) in monkeys that had been trained to grasp objects using two different tools: "normal pliers" and "reverse pliers" (figure 43.2*A*). These tools require opposite movements to grasp an object: With normal pliers, the hand has to be first opened and then closed, as when grasping is executed with the bare hand; with reverse pliers, the hand has to be first closed and then opened. The use of the two tools enabled the researchers to dissociate the neural activity related to hand movement from that related to the goal of the motor act.

All tested neurons in area F5 and half of neurons recorded from the primary motor cortex discharged in relation to the accomplishment of the goal of grasping—when the tool closed on the object—regardless of whether in this phase the hand opened or closed, that is, regardless of the movements that were employed to accomplish the goal (figure 43.2*B*). These data indicate that goal coding is at the basis of the organization of grasping in area F5 and also in the primary motor cortex, although to a minor extent. Goal coding is therefore not an abstract, merely mentalist and experience-independent property, but a distinctive functional feature upon which the cortical motor system is organized.

An organization based on goal-directed hand motor acts is also present in AIP and in PFG (Murata et al., 2000; Fogassi et al., 2005; Rozzi, Ferrari, Bonini, Rizzolatti, & Fogassi, 2008). In both these areas, there are neurons that code specific motor acts and specific types of grips. At present, there are no systematic studies in which the functional properties of these areas have been compared with F5. As far as one can deduce from the available studies, there are strong similarities between neurons with motor properties in these areas (Raos et al., 2006). Taken together, these data indicate that the rostral part of the inferior parietal lobule is functionally part of the motor system in the same way as the premotor areas that belong to it.

Recent data by Fogassi and colleagues (2005) showed that the discharge of IPL neurons coding grasping is influenced by the action in which grasping is embedded. In this study, PFG grasping neurons were tested in two main conditions. In one, the monkey reached for and grasped a piece of food located in front of it and brought the food to its mouth. In the other, the monkey reached for and grasped an object and placed it into a container (figure 43.3*A*). The results showed that the majority of the recorded neurons discharged with a different intensity depending on the final goal of the action (eating or placing) in which grasping was embedded ("action-constrained" neurons) (figure 43.3*B*). A series of controls for grasping force, kinematics of reaching movements, and type of stimuli showed that neuron selectivity was not due to these factors. Thus, the differential discharge of these grasping neurons appeared to reflect the goal of the action of which the motor act was part. A similar organization has recently been reported for area F5 (Fogassi et al., 2007).

These neural properties suggest that most neurons of the PFG-F5 circuit code individual motor acts (e.g., grasping) in prewired chains, each of them coding a specific action (e.g., eating). This organization is very appropriate for providing fluidity to action execution, because each neuron not only codes a specific motor act, but, being embedded into a specific action, is also linked with neurons that code the next motor acts and possibly facilitates them. In favor of a model that assumes a facilitatory interaction between neurons forming a given chain is the organization of the receptive fields of IPL. For example, there are IPL neurons that respond to passive stimulation of the hand, flexion of the forearm, and discharge during mouth grasping (Yokochi, Tanaka, Kumashiro, & Iriki, 2003; Rozzi et al., 2008). These data support the existence of chains of neurons that code specific actions such as that of bringing food to the mouth.

Functional Properties of MNs As has already been stated, mirror neurons are a distinct class of visuomotor neurons that discharge both when individuals perform a specific motor act *and* when they observe the same motor act done by another individual (figure 43.4). Among the motor acts that they code both visually and motorically, the most

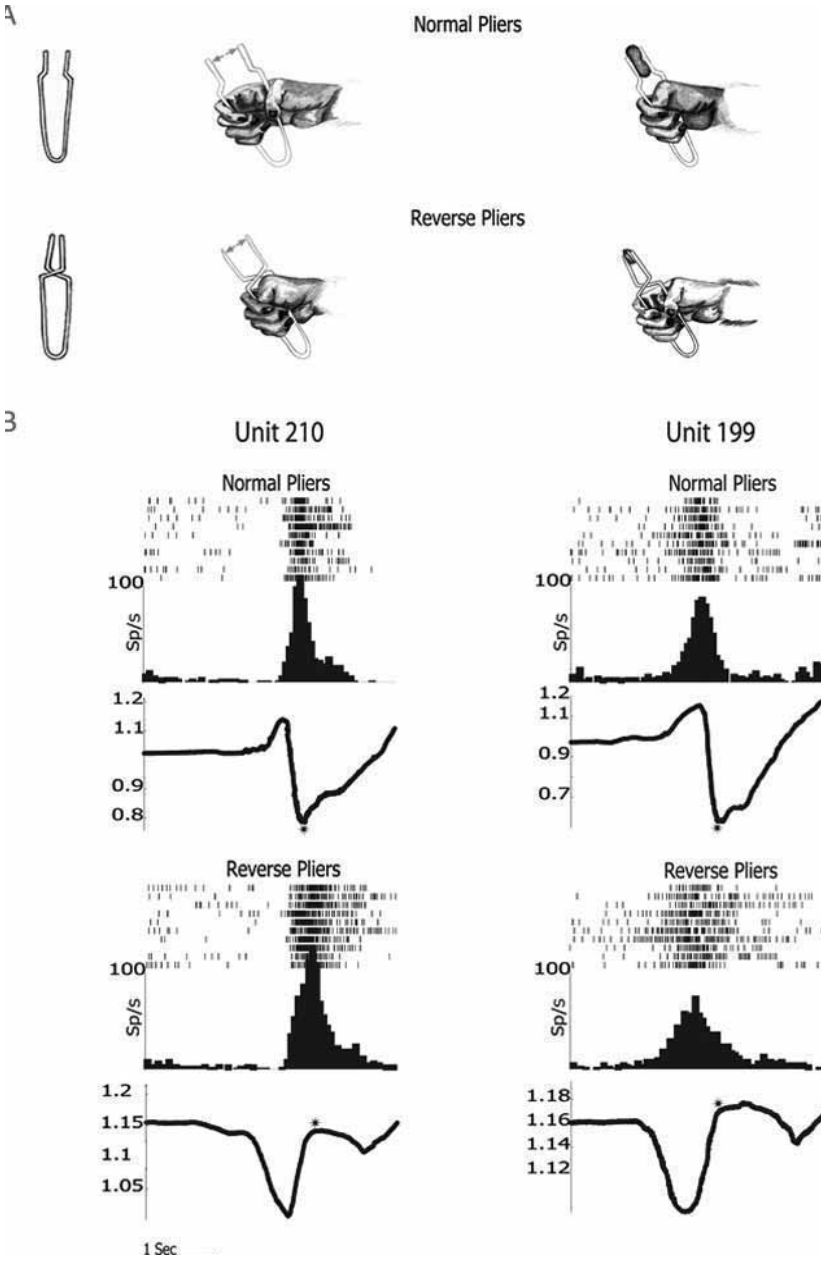


FIGURE 43.2 Examples of F5 neurons active during execution of grasping with normal and reverse pliers. (*A*) Illustration of the experimental paradigm. To grasp the object with the normal pliers (upper part), the monkey has to close its hand, while to grasp the object with the reverse pliers (lower part), the monkey has to open its hand. (*B*) Two neurons recorded in area F5. Rasters and histograms are aligned with the end of the grasping closure phase (asterisk). The traces below each histogram indicate the hand position, recorded with a potentiometer, expressed as a function of the distance between the pliers handles. When the trace goes down, the hand closes; when the trace goes up, it opens. The values on the vertical axes indicate the voltage change measured with the potentiometer. (Modified from Umiltà et al., 2008.)

represented are grasping, holding, manipulating, and tearing. Unlike another category of visuomotor neurons that are present in area F5 ("canonical neurons") (Murata et al., 1997; Raos et al., 2006), they do not fire in response to simple presentation of objects, including food. The observation of intransitive motor acts, including mimed motor acts, is also ineffective.

Mirror neurons show a close relationship between their visual and motor responses. Using as classification criterion the congruence between the executed and observed motor acts that are effective in triggering them, mirror neurons have been subdivided into two broad classes: strictly congruent and broadly congruent neurons (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). They are defined as *strictly congruent* when

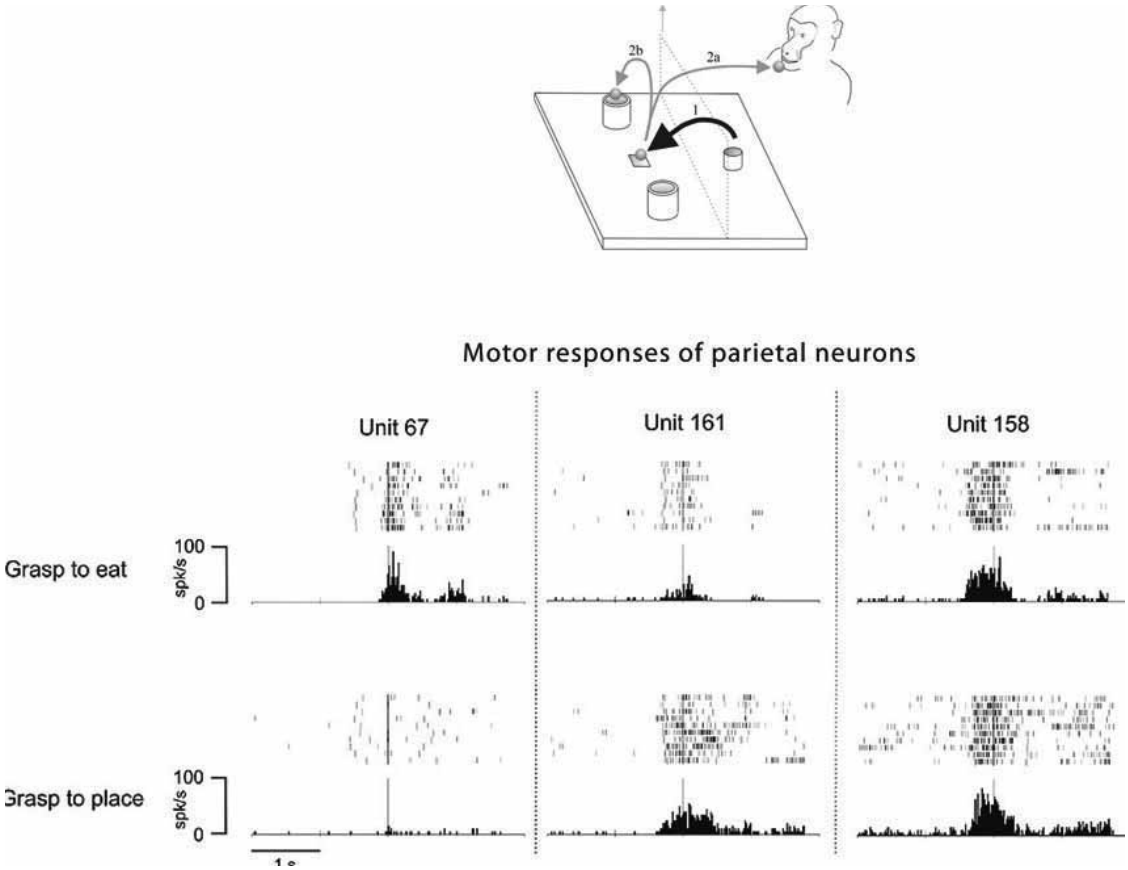


FIGURE 43.3 Examples of the activity of parietal motor neurons during execution of two different actions. (*A*) Apparatus and paradigm used for the motor task. In one condition (grasping for eating), the monkey reached for and grasped a piece of food located on a plane in front of it (1) and brought the food to its mouth (2a). In another condition (grasping for placing), the monkey reached for and grasped an object located in front of it (1) and placed the object into a container (2b). In the first condition, the monkey ate the food that it had brought to the mouth; in the second condition, the monkey was rewarded after correct accomplishment of the task. (*B*) Activity of three IPL neurons during grasping in the two experimental conditions. Unit 67 discharges were stronger during grasping to eat than during grasping to place, Unit 161 discharges were stronger during grasping to place. Unit 158 did not show any difference in discharge between the two conditions. Rasters and histograms are aligned with the moment when the monkey touched the object or food to be grasped. Red bars: Monkey releases the hand from the starting position. Green bars: Monkey touches the container. Abscissa: Time, bin = 20 ms; Ordinate: Discharge frequency in spikes per second. (Modified from Fogassi et al., 2005.) (See color plate 55.)

the observed and executed effective motor acts are identical in terms of goal (e.g., grasping) and in terms of the way in which that goal is achieved (e.g., precision grip). In contrast, mirror neurons are defined as *broadly congruent* when there is a similarity, but not identity, between the observed and executed effective motor acts. Among the different types of broadly congruent neurons, the most common is constituted of neurons that become active during the execution of a specific motor act made by the monkey (e.g., grasping, holding, or manipulating) but visually respond to more than one motor act (e.g., manipulation and grasping).

In the first studies on mirror neurons, it was reported that these neurons do not discharge during the observation of goal-directed actions done by using tools (Gallese et al., 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Subsequently, however, it was shown that following a relatively long period during which monkeys observed the experimenters performing actions using tools, some mirror neurons respond, although weakly, also to this type of action (Rizzolatti & Arbib, 1998). More recently, Ferrari, Rozzi, and Fogassi (2005) reported that in a specific ventral sector of F5, there are neurons that discharge very vigorously to the observation of tool use (e.g., a stick or a pair of pliers). It is not clear whether these neurons, like those previously observed, derived this property because of prolonged action observation.

The most widely accepted hypothesis on the functional role of mirror neurons is that they play a role in understanding the goal of the observed motor acts (Rizzolatti et al., 2000). The proposed mechanism is the following: individuals know the outcome of their motor acts. Thus, when the mirror neurons of an observing individual, which code a

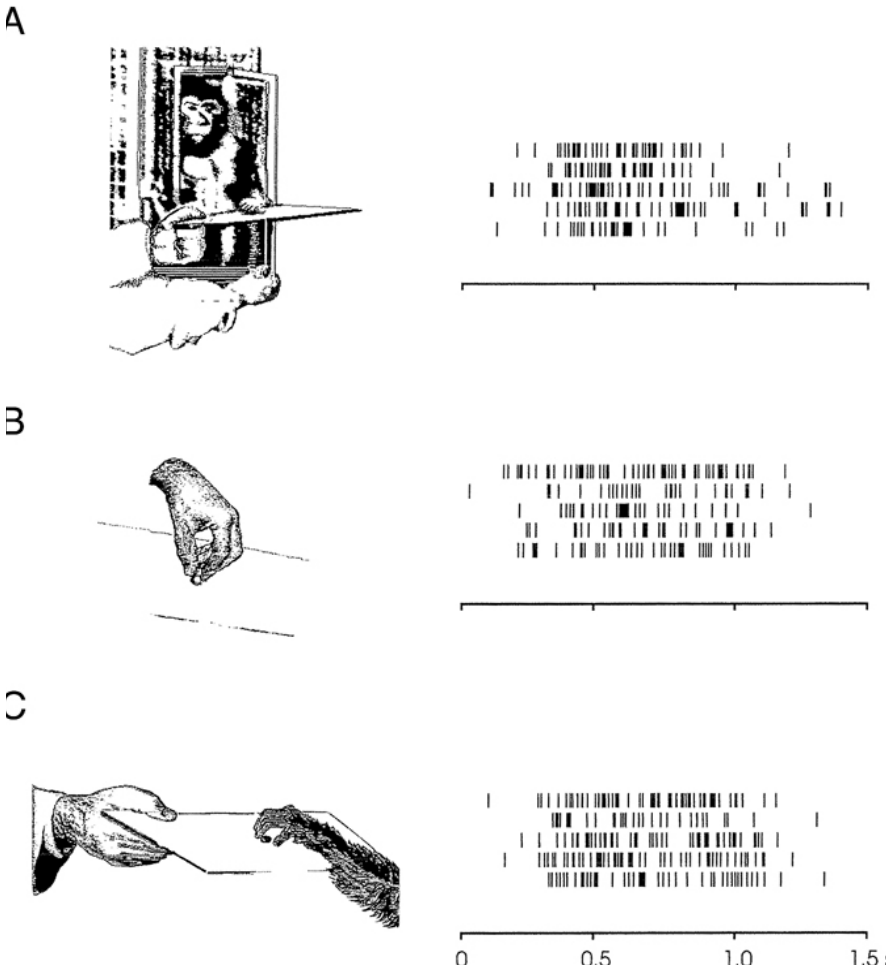


FIGURE 43.4 Example of the activity of an F5 mirror neuron during the observation of a grasping movement performed by another monkey (*A*) or an experimenter (*B*) and during execution of grasping by the recorded monkey (*C*). (Modified from Rizzolatti et al., 1996.)

given motor act (e.g., grasping), discharge in response to the observation of that motor act (grasping) done by another individual, the observer understands its goal, because that discharge corresponds to the one that occurs when the observer wants to achieve the same goal.

To provide evidence in favor of the view that mirror neurons play a role in understanding motor acts done by others, neurons' responses were investigated when the monkeys could comprehend the goal of a motor act without actually seeing it. If mirror neurons truly mediate understanding, their activity should reflect the meaning of the motor act rather than its visual features. Two series of experiments were carried out for this purpose.

The first series tested whether mirror neurons could recognize motor acts merely from their sounds (Kohler et al., 2002). The activity of mirror neurons was recorded while a monkey was observing a motor act, such as ripping a piece of paper or breaking a peanut shell, that is normally accompanied by a distinctive sound. Then the monkey was presented with the sound alone. It was found that many mirror neurons that had responded to visual observation of acts accompanied by sounds also responded to the sound alone. These neurons were named "audiovisual" mirror neurons.

In the second series of experiments, the researchers hypothesized that if mirror neurons are involved in understanding a motor act, they should also discharge when the monkey does not actually see the motor act but has sufficient clues to create a mental representation of it. Therefore, F5 mirror neurons were tested in two conditions. In one, the monkey was shown a fully visible motor act directed toward an object ("full vision" condition). In the other, the monkey saw the same act but with its final critical part hidden ("hidden" condition) (Umiltà et al., 2001). The results showed that more than half of F5 mirror neurons also discharged in the hidden condition. An example is shown in figure 43.5.

These experiments strongly support the notion that the activity of mirror neurons underpins the understanding of motor acts. Even when the motor act comprehension is possible on a nonvisual basis, such as via sound or nonlinguistic

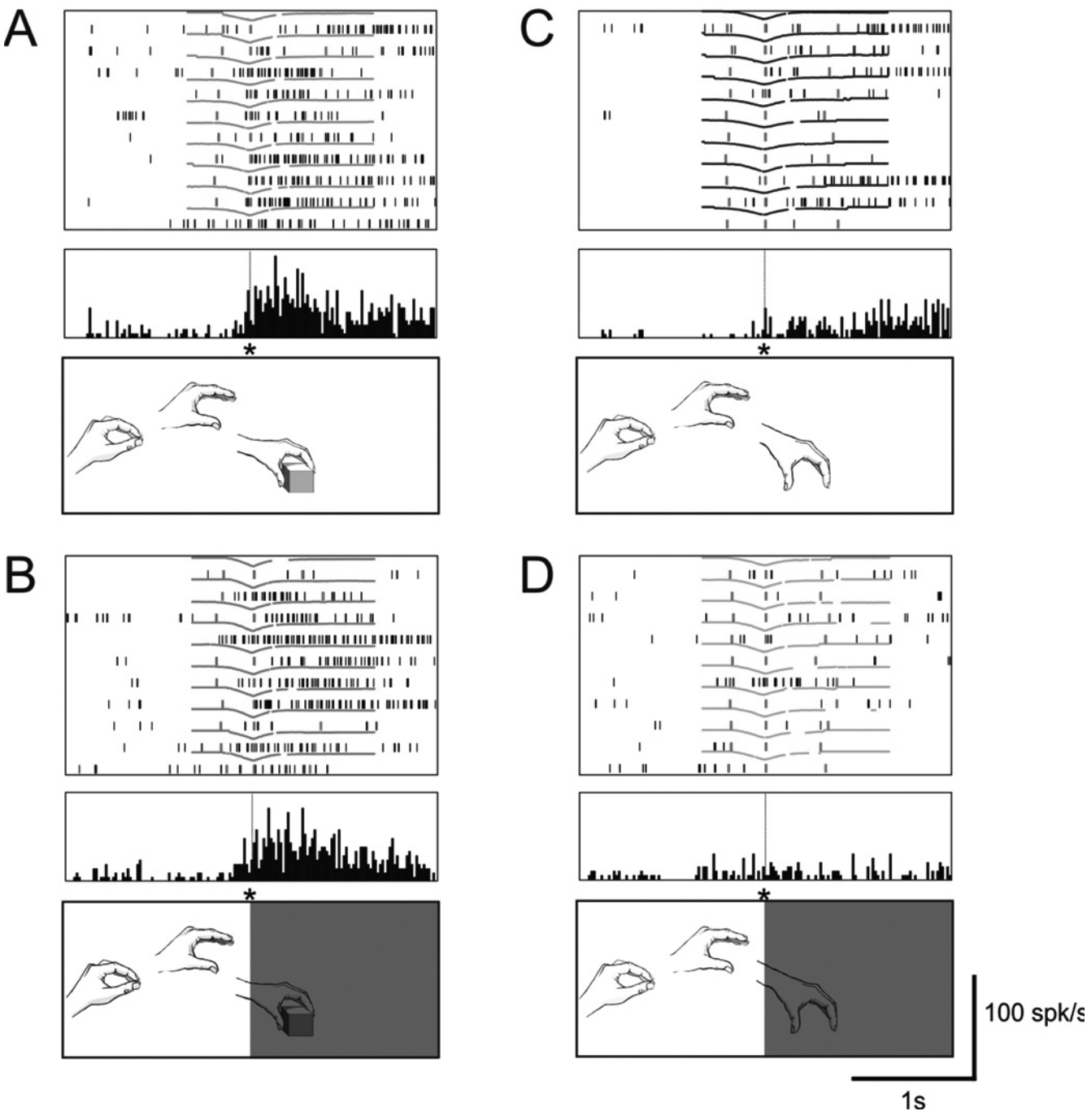


FIGURE 43.5 Example of a mirror neuron responding during observation of grasping in both full vision and "hidden" condition. (*A* and *C*): Observation of goal-directed or mimed grasping, respectively, in full vision. (*B* and *D*): Observation of goal-directed or mimed grasping, respectively, in the hidden condition. In every panel, from top to bottom, rasters and histogram and the schematic drawing of the experimenter motor act are shown. The gray frame in conditions *B* and *D* represents a screen interleaved between the monkey and the experimenter hand in the two hidden conditions. The asterisk indicates the location of a stationary marker that was attached at the level of the crossing point where the experimenter's hand disappeared behind the screen in the hidden conditions. The colored line above each raster represents the kinematics of the experimenter's hand movement; the downward deflection of the line means that the hand is approaching the stationary marker (the minimum corresponding to the moment in which the hand is closest to the marker). Histograms bin width = 20 ms. (Modified from Umiltà et al., 2001.) (See color plate 56.)

mental representation, mirror neurons equally discharge, signaling the goal of the motor act.

Early studies on mirror neurons examined the dorsalmost sector of F5, where hand motor acts are mostly represented. Recently, a study was carried out on the properties of neurons located in the most ventral part of F5, where neuron activity is mostly related to mouth actions (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003). The results showed that two classes of mouth mirror neurons could be distinguished: ingestive and communicative mirror neurons. Ingestive mirror neurons, which represent the majority of mouth mirror neurons, respond to the observation of motor acts related to ingestive functions (e.g., grasping food with the mouth). Virtually all of them show a good correspondence between the effective observed and the effective executed motor act. More intriguing are the properties of the communicative mirror neurons. For them, the most effective observed motor act is a communicative gesture, such as lip smacking. However, most of them strongly discharge also when the monkey actively performs an ingestive motor act. The presence of a motor response during both communicative and ingestive motor acts is rather intriguing. However, it could be explained by ethological observations suggesting that in evolution, monkeys' communicative gestures derived, at least in part, from ingestive motor acts (Van Hoof, 1962, 1967; Maestripieri, 1996).

Intention understanding

Before we discuss the role of mirror neurons in intention understanding, it is important to define the terms *motor act* and *motor action*. *Motor act* describes a movement or, most commonly, a series of movements performed to achieve a goal (e.g., grasping an object). *Motor action* describes a series of motor acts (e.g., reaching, grasping, bringing to the mouth) that allow individuals to fulfill their intention (e.g., eating). When an individual observes a motor act, he or she understands the *what* of the motor act (e.g., grasping an object) but typically is also able to make inferences about *why* the motor act is being performed (e.g., grasping for eating), that is, the intention behind the action of which that motor act is part.

As was described above, grasping neurons in both parietal and premotor cortex discharge with a different intensity according to the final goal of the action in which the grasping act is embedded (action-constrained neurons). Further experiments investigated whether action-constrained neurons also had mirror properties and whether their visual response during grasping observation was influenced by the action goal in which grasping was embedded (Fogassi et al., 2005, 2007). To this purpose, neurons were tested in the same two conditions that were used for studying their motor properties. Instead of grasping objects, monkeys observed the experimenter performing the two actions (grasping for eating and grasping for placing).

The results showed that the majority of mirror neurons in the two areas were differently activated when the observed motor act belonged to one action or another. Examples are shown in figure 43.6.

What could be the explanation of this neuron behavior? It is very likely that when an action-constrained grasping neuron is activated by the observation of a grasping motor act inserted into its motor action, its discharge triggers the whole motor chain of the observer underpinning the same action. In this way, the observer activates an internal motor representation of the action that the observed agent intends to do. Thanks to this mechanism, the observer understands the observed agent's intention.

One may ask how action observation can activate the appropriate motor chain when the monkey actually sees only the first motor act of it. A systematic study of this problem has not been done. It is clear, however, from the grasping neuron behavior that an important factor in determining the neuron discharge is the type of stimulus with which the agent interacts. Food, for example, tends to activate eating chains as soon as the monkey sees the experimenter grasping the food. Another factor is the statistical probability of a given action. Thus, for example, in a block of trials in which grasping is always followed by placing, grasping neurons that are tuned for placing become active. It is interesting to note that in such a block of trials, if food, rather than an object, is grasped and placed into a container, grasping-to-eat neurons fire initially, then they stop firing while grasping-to-place neurons become active.

Mirror-like Activity

Recent data suggest that neurons in dorsal premotor and primary motor cortex discharge during execution and observation of trained arm movements directed toward a target.

In one study (Cisek & Kalaska, 2005), monkeys were trained to move a cursor from a central position to a peripheral position on a screen defined by a color cue. The recorded neurons discharged both when the monkey performed the learned task and when the monkey, being still, observed another party moving the cursor in the correct direction. The discharge typically occurred at the presentation of the target and increased with the cursor movement. Unlike mirror neurons, these neurons did not require the observation of an effector-object interaction. One may postulate, however, that the cursor was an abstract substitute for a moving hand and that the occurrence of the stimulus evoked the mental representation of the hand movement.

In another study (Tkach, Reimer, & Hatsopoulos, 2007) monkeys were trained to move repetitively a cursor to targets

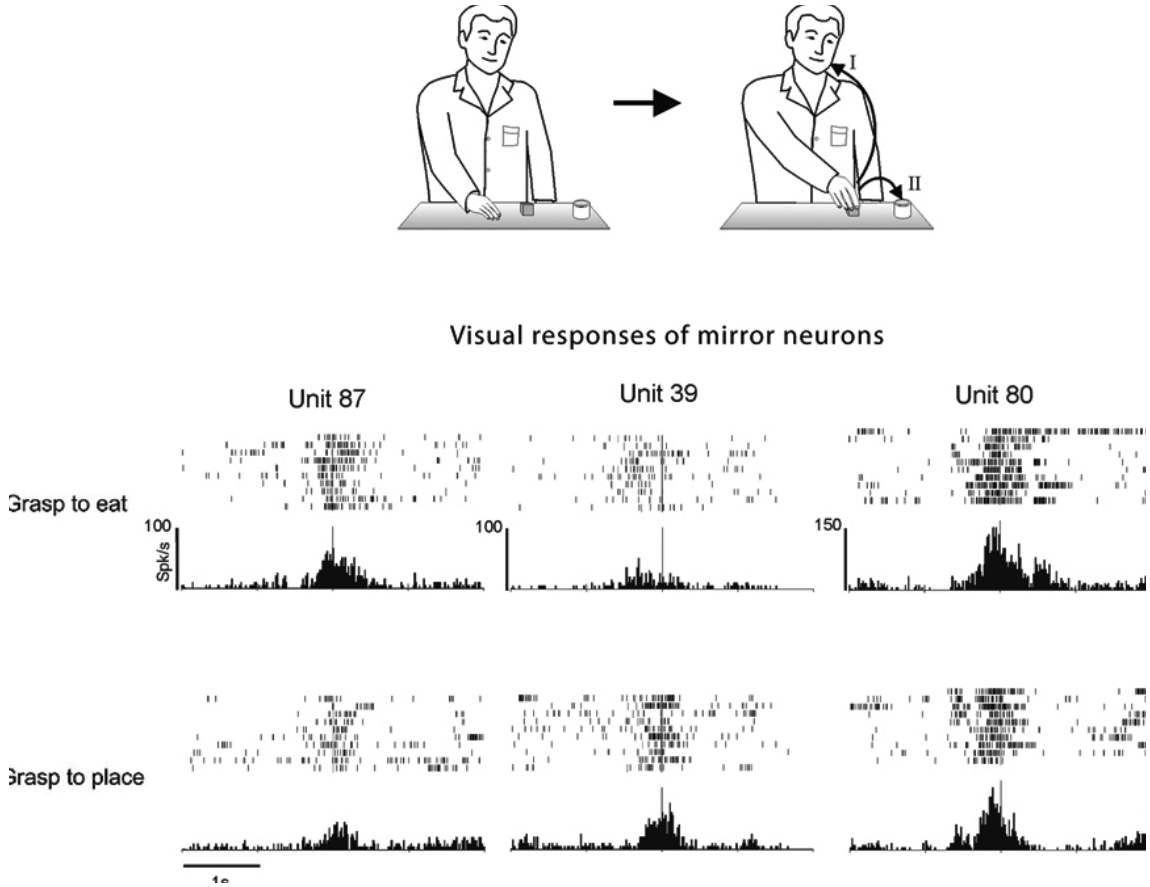


FIGURE 43.6 Examples of visual responses of IPL mirror neurons during the observation of grasping-to-eat and grasping-to-place conditions performed by an experimenter. (*A*) The paradigm is similar to that used for the motor task shown in figure 43.3, but in this case, the two conditions are performed by the experimenter in front of the monkey, which is simply observing the scene. (*B*) Activity of three mirror neurons during observation of grasping in the two experimental conditions. Unit 87 discharges are stronger during observation of grasping to eat than during observation of grasping to place, Unit 39 discharges are stronger during observation of grasping to place. Unit 80 did not show any difference in discharge between the two conditions. Rasters and histograms are aligned with the moment when the experimenter touched the object or food to be grasped. (Modified from Fogassi et al., 2005.) (See color plate 57.)

that appeared at random locations. The experiment consisted of two phases: active movement and observation. In the active movement phase, the monkey controlled the cursor, while in the observation phase, the monkey observed the replayed movements generated in the active phase. The observation phase had three conditions. In the first, both the cursor and the targets were visible; in the second, the monkey saw only the replayed targets; in the third, the monkey saw only the moving cursor but not the targets. The results showed that passive observation of the task determined a neural discharge similar to that found during task execution. The observation of the cursor without targets or of the targets without cursor gave either no responses or responses that were weaker than those found during the observation of both cursor and targets. The authors concluded that the most likely explanation of their findings is that the observation of the movements determined a covert generation of a motor command.

Mirror system in humans

ANATOMY OF THE MIRROR SYSTEM A large number of brain imaging studies showed that parietal and frontal areas that activate during motor acts execution are also active when an individual observes similar motor acts done by others (see Rizzolatti & Craighero, 2004). Most of these studies concerned observation of object-directed *grasping* movements. The regions that are activated in these studies form the grasping human mirror system. The two main nodes of this system are the inferior parietal lobule (IPL) and the ventral premotor cortex (PMv) plus the caudal part of the inferior frontal gyrus (IFG), roughly corresponding to its *pars opercularis*. The localization of human grasping mirror system corresponds to that of the homologous mirror system in the monkey (figure 43.7).

Several experiments addressed the issue of how observed motor acts performed by different effectors are organized in

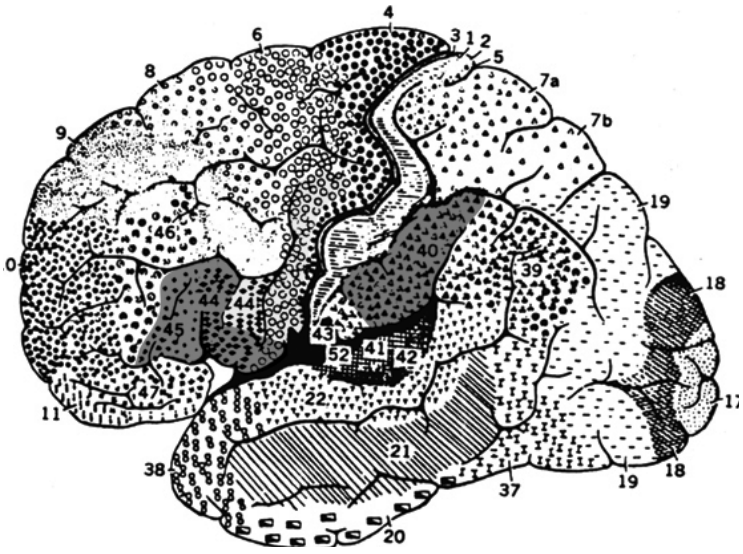


FIGURE 43.7 Lateral view of the human cortex showing the frontal (yellow and blue) and parietal (red) regions constituting the core of the grasping mirror neuron system in humans. Numbers and symbols indicate the different cytoarchitectonic areas according to the parcellation of Brodmann (1909). (See color plate 58.)

the human mirror system by presenting videos of motor acts performed with leg, hand, and mouth (Buccino et al., 2001; Sakreida, Schubotz, Wolfensteller, & von Cramon, 2005; Shmuelof & Zohary, 2006; Wheaton, Carpenter, Mizelle, & Forrester, 2008) or using point-light displays of biological motion of different body parts (Saygin, Wilson, Hagler, Bates, & Sereno, 2004; Ulloa & Pineda, 2007).

As far as the premotor cortex is concerned, the results showed that the observed leg motor acts are represented more dorsally in the ventral premotor cortex (PMv) extending across the superior frontal sulcus into the dorsal premotor cortex (PMd), and the hand motor acts are represented in an intermediate position in PMv, while the mouth motor acts are represented ventrally, extending into the IFG. There was considerable overlap between adjacent representations. While the goal of the observed motor acts in these studies was achieved mostly by distal movements, a recent study investigated the organization of reaching movements, that is, the transport phase of the hand to a particular location in space, eliminating the contribution of grasping movements (Filimon, Nelson, Hagler, & Sereno, 2007). It was found that in both observation and execution, the sector of premotor cortex that was activated was located in the cortex of the superior frontal gyrus (SFG), that is, in PMd. Thus, it appears that observation of motor acts focused on the distal part of the effector activates PMv, while when the focus is on the proximal part, activation mostly concerns PMd.

The activation pattern in the parietal lobe is rather complex. The observation of goal-directed motor acts in which the focus was on distal movements showed activation of the rostral part of the cortex inside and around the intraparietal sulcus, extending into the convexity of IPL for mouth motor acts, activation of the caudal part of the same cortex but extending into the superior parietal lobule for the leg motor acts, and activation of an intermediate sector for the hand motor acts (Buccino et al., 2001). In the experiment (Filimon et al., 2007) in which the focus was on observation of the transport phase (reaching movement), the activation was located more dorsally, in the superior parietal lobule extending toward the dorsal bank of the IPS.

In a recent study (Jastorff, Rizzolatti, & Orban, 2007), video clips showing four distal motor acts (grasping, dragging, dropping, and pushing), each performed by using three different effectors (foot, hand, and mouth), were presented to volunteers. The results showed that while in PMv, the activations determined by the observed motor acts were clustered according to the effector used, independently of their positive (grasping and dragging) or negative (dropping and pushing) behavioral valence, in the parietal cortex, the organization followed another principle: The observed motor acts were found to be clustered according to their valence, regardless of whether they were done with the mouth, hand, or foot. The most activated region corresponded to putative human AIP, extending ventrally to the inferior parietal lobule and dorsally to the superior parietal lobule. Motor acts with negative valence were represented dorsally, while those with positive valence ventrally. It can be hypothesized that this parietal organization, by generalizing the motor act valence across effectors, allows a unified understanding of the observed behavior.

In addition to an organization based on the valence of the motor act, the parietal lobe activation also showed a coarse effector-based organization. The strongest activations for foot motor acts were located dorsally, and those for mouth

motor acts were located ventrally, well below AIP. Activations for hand motor acts were the strongest in the center of the responsive region, which also responds to foot acts. It has been suggested (Jastorff et al., 2007) that the valence-related organization is based on a motor scaffold and that the representation of the observed motor acts that are typically performed with the hand becomes active also when they are performed with other effectors.

Another issue that has been recently addressed (Gazzola, Rizzolatti, Wicker, & Keysers, 2007; Orban et al., 2006) is whether the observation of tool use or robotic arms activates the same circuit that becomes active during the observation of motor acts done with natural effectors. The results of these studies showed that the basic parieto-premotor circuit that becomes active during the observation of hand grasping is also active during the observation of tool actions. In addition, however, it was shown (Orban et al., 2006) that the observation of actions performed with tools activates a specific region in the inferior parietal lobule, corresponding to the rostral inferior part of the supramarginal gyrus. The two parietal regions that are activated by the observation of tools and robotic arms could underlie two different ways in which tool use is understood. The sector around the intraparietal sulcus could mediate an association between a tool and the tool use outcome without an understanding of tool functioning. In contrast, the rostral supramarginal gyrus could be involved in the uniquely human capacity of understanding the tool use in terms of its functioning. It is interesting to note that the rostral supra-marginal gyrus is the part of the inferior parietal lobule that is most frequently damaged in patients with ideomotor apraxia (see Leiguarda & Marsden, 2000; Wheaton & Hallett, 2007).

Plasticity of the Mirror System Is the mirror system modulated by motor experience? There is clear evidence that the observation of motor acts that are richly represented in the observer's motor repertoire determines a stronger activation of the mirror system than does the observation of novel motor behaviors (Casile & Giese, 2006). In particular, in a functional magnetic resonance imaging (fMRI) study, Calvo-Merino, Glaser, Grezes, Passingham, and Haggard (2005) demonstrated that the observation of actions performed by others results in different cortical activations depending upon the specific motor competence of the tested individuals. Participants, who included classical dancers, dancers of Capoeira, and people who had never taken a dancing class, were shown a video of Capoeira dance steps. The sight of the dance steps of Capoeira caused a greater activation of the mirror system in the Capoeira dancers than in either the classical dancers or the beginners. Conversely, a video showing classical dance steps resulted in a much stronger activation of the classical dancers' mirror system compared to those of the Capoeira dancers and the beginners.

In a further experiment, the same researchers (Calvo-Merino, Grezes, Glaser, Passingham, & Haggard, 2006) tried to understand whether the differences in the activation found in the previous experiment were due to motor or visual familiarity with the observed movements. The results showed that the mirror system was activated more strongly by the sight of the dance steps executed by the dancers of the same sex of the observer, indicating, therefore, that the activation was regulated by motor practice and not by visual experience, given the fact that the latter was the same for both sexes.

The data by Calvo-Merino and colleagues (2005, 2006) were extended by Cross, Hamilton, and Grafton (2006) in a study in which expert dancers learned and rehearsed novel, complex whole-body dance sequences for five weeks. Brain activity was recorded weekly by fMRI as dancers observed and imagined performing different movement sequences. Half of these sequences were rehearsed, and half were unpracticed control movements. Critically, activation of the mirror system was modulated as a function of dancers' motor experience.

These data show that the mirror system codes the observed actions by mapping them onto "corresponding" motor representations of the observer. But how would the mirror system respond to the observation of hand actions if the observer never had hands or arms? Two aplasic individuals, born without arms or hands, were scanned while they observed hand motor acts (Gazzola, van der Worp, et al., 2007). The results showed activations in the parietofrontal circuit of aplasic individuals while they watched hand motor acts. This finding demonstrates the brain's capacity to mirror acts that deviate from the typical motor organization by recruiting brain cortical representations involved in the execution of motor acts that achieve corresponding goals by using different effectors.

The Mirror System and Intention Understanding in Humans Recent experiments showed that besides understanding of motor acts, the mirror system is also involved in understanding the *intention* behind the observed motor acts.

Evidence in this sense has been provided by an fMRI study (Iacoboni et al., 2005). In this study, there were three conditions. In the first one (called "context"), participants saw a scene with objects (a teapot, a mug, a plate with some food on it) arranged as if a person was ready to have breakfast or arranged as if a person had just finished having breakfast; in the second condition (called "action"), participants were shown a hand that grasped a mug without any context; in the third (called "intention"), participants saw the same hand motor act within the two different contexts. The

context suggested the intention of the agent, that is, grasping the cup for drinking or grasping it for cleaning.

The results showed that in both action and intention conditions, there was an activation of the mirror system. The comparison between intention and action conditions showed that the understanding of the intention of the agent determined a marked increase in activity of the right IFG.

Interestingly, the observation of grasping of the cup to drink produced a stronger activation than did the observation of grasping done to clean. This result is similar to findings in monkeys (see above) showing that the number of neurons that code grasping for bringing to the mouth largely exceeds the number of neurons that code grasping for putting an object into a container.

In another fMRI study, based on repetition suppression paradigm, participants were instructed to observe repeated movies showing the same action outcome (such as opening or closing a box) achieved by using the same or different kinematics. The results showed that the right inferior parietal and right inferior frontal cortex responses were suppressed by the observation of the same action outcome, independent of the means used to achieve it. This finding has been interpreted as evidence of the involvement of the mirror system in intention understanding (de Hamilton & Grafton, 2008).

In conclusion, these data show that the intentions behind the actions—at least of basic actions—of others can be recognized through the mirror mechanism. This does not imply, of course, that other, more cognitive ways of "reading minds" do not exist (see Frith & Frith, 2008). However, there is little doubt that the mirror mechanism is one of the most basic and possibly the most basic mechanism for intention understanding.

More recently, an fMRI study investigated the neural basis of human capacity to differentiate between actions that reflected the intention of the agent (intended actions) and actions that did not reflect it (nonintended actions). Participants were shown video clips of a variety of actions done with different effectors, each in a double version: one in which the actor achieved the purpose of his or her action (e.g., pouring the wine) and one in which the actor performed a similar action but failed to reach the goal because of a motor slip or a clumsy movement (e.g., spilling the wine) (Buccino et al., 2007). The results showed that both types of actions activated the mirror system. The direct contrast nonintended versus intended actions showed activation in the right temporoparietal junction, left supramarginal gyrus, and mesial prefrontal cortex. The converse contrast did not show any activation. It was concluded that the capacity to understand when an action is nonintended is based on the activation of attention areas signaling unexpected events in spatial and temporal domains (Corbetta & Shulman, 2002; Coull, 2004; Mitchell, 2008). These results indicate that when an individual observes an unexpected motor act, such as a motor slip, his cortical machinery, besides signaling the observed motor act, also signals the strangeness of the motor act outcome.

The Mirror System, Motor Cognition, and Autism

Autistic children display a striking inability to relate themselves to people in ordinary ways. According to Kanner (1943), this represents the fundamental feature of autism. However, in the same seminal paper, Kanner reported that "almost all mothers [. . .] recalled their astonishment at the children's *failure to assume at any time an anticipatory posture* (Kanner's italics) preparatory to being picked up." Unlike typically developing children, autistic children use motor strategies that basically rely on feedback information rather than on feedforward modes of control. Such motor disturbance prevents autistic children from adopting anticipatory postural adjustments (Schmitz, Martineau, Barthélemy, & Assaiante, 2003).

The theoretical relevance of these findings has been clarified by a recent electromyelographic (EMG) study (Cattaneo et al., 2008) showing that high-functioning autistic children are unable to organize their own motor acts in intentional motor chains as typically developing children do. Participants in this study were typically developed (TD) and high-functioning autistic children who were required both to execute and to observe two different actions: grasping with the right hand a food item placed on a plate, bringing it into the mouth, and eating it or grasping a piece of paper placed on the same plate and putting it into a box (figure 43.8*A*). During the execution and observation conditions of both actions, the activity of the mouth-opening mylohyoid muscle (MH) of the participants was recorded by using EMG surface electrodes.

The results showed that during the execution and observation of the eating action, a sharp increase of MH activity was recorded in TD children, starting well before the food was grasped. No increase of MH activity was present during the observation of the placing action. This means that one of the muscles that are instrumental to accomplish the action final goal (opening the mouth to eat a piece of food) is already activated during the initial phases of the action. The motor system anticipates the consequences of the action final goal (to eat), thus directly representing the action intention, both when the action is executed and when the action is observed being done by others. In contrast with TD children, high-functioning autistic children showed a much later activation of the MH muscle during eating action execution and *no activation at all* during eating action observation (figures 43.8*B* and 43.8*C*).

These results reveal that children with autism are impaired in chaining sequential motor acts within a reaching-to-grasp-to-eat intentional action, a mechanism

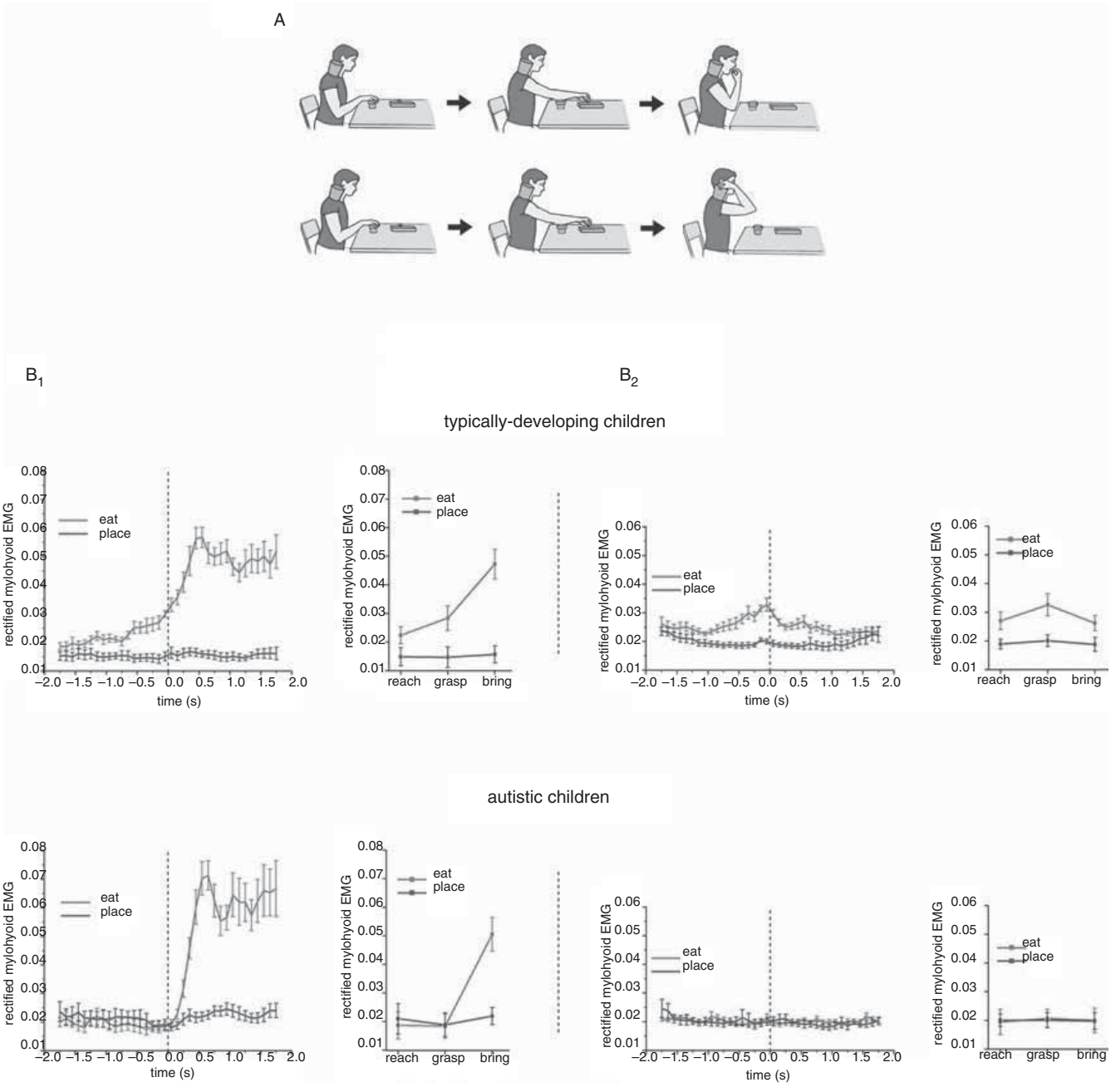


FIGURE 43.8 Differential activation of a mouth-opening muscle during execution and observation of two actions in typically developing and autistic children. (*A*) Schematic representation of the two actions executed and observed by the two groups of subjects. *Upper part*: The individual reaches for and grasps a piece of food located on a touch-sensitive plate, brings it to the mouth, and eats it. *Lower part*: The individual reaches for and grasps a piece of paper located on the same plate and puts it into a container placed on the shoulder. (B_1) *Left*: Time course of the EMG activity of the mylohyoid muscle during execution of grasping for eating (red) and grasping for placing (blue). Vertical bars indicate the standard error. The curves are aligned (dashed vertical line) with the moment in which the object is lifted from the touch-sensitive plate. *Right*: Mean EMG activity of the same muscle in three epochs of the two actions. Vertical bars indicate 95% confidence intervals. (B_2) *Left*: Time course of the EMG activity of the mylohyoid muscle during observation of grasping for eating (red) and grasping for placing (blue). Other conventions as in B_1. *Right*: Mean EMG activity of the same muscle in three epochs of the two observed actions. Other conventions as in B_1. (Modified from Cattaneo et al., 2008.) (See color plate 59.)

that most likely reflects the chained organization of the parietal cortex described in the monkey (see above) (Fogassi et al., 2005). This impairment is mirrored in the action observation condition and most likely accounts for the difficulty these children have in *directly understanding* the intention of the observed action when executed by others.

Other recent studies have documented a deep impairment of the core mechanisms of motor cognition in children with autism. Two recent studies show that autistic individuals might be suffering from a dysfunction of their mirror system. Theoret and colleagues (2005) showed that, again unlike healthy controls, children with autism did not show TMS-induced hand muscle facilitation during hand action observation. Oberman and colleagues (2005) showed that children with autism, unlike healthy controls, did not show mu frequency suppression over the sensorimotor cortex during action observation.

Hence, converging evidence from a variety of studies suggests that some of the social cognitive impairments manifested by autistic individuals could be rooted in their incapacity to organize and directly grasp the intrinsic goal-related organization of motor behavior, because of a dysfunctional mirror system.

REFERENCES

Alexander, G. E., & Crutcher, M. D. (1990). Neural representations of the target (goal) of visually guided arm movements in three motor areas of the monkey. *J. Neurophysiol.*, *64*, 164–178.

Belmalih, A., Borra, E., Contini, M., Gerbella, M., Rozzi, S., & Luppino, G. (2009). Multimodal architectonic subdivision of the rostral part (area F5) of the macaque ventral premotor cortex. *J. Comp. Neurol., 512*(2), 183–217.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Curr. Biol.*, *17*, 2117–2121.

Brodmann, K. (1909). *Vergleichende Lokalisationlehre der Grosshirnrinde*. Leipzig: Barth.

Buccino, G., Baumgaertner, A., Colle, L., Buechel, C., Rizzolatti, G., & Binkofski, F. (2007). The neural basis for understanding non-intended actions. *NeuroImage*, *36*(Suppl 2), T119–127.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *Eur. J. Neurosci.*, *13*, 400–404.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An FMRI study with expert dancers. *Cereb. Cortex*, *15*, 1243–1249.

Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr. Biol.*, *16*, 1905–1910.

Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Curr. Biol.*, *16*, 69–74.

Cattaneo, L., Fabbi-Destro, M., Boria, S., Pieraccini, C., Monti, A., Cossu, G., & Rizzolatti, G. (2008). Impairment of action chains in autism and its possible role in intention understanding. *Proc. Natl. Acad. Sci. USA*, *104*, 17825–17830.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, *45*, 801–814.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.*, *3*, 201–215.

Coull, J. T. (2004). fMRI studies of temporal attention: Allocating attention within, or towards, time. *Brain Res. Cogn. Brain Res.*, *21*, 216–226.

Cross, E. S., Hamilton, A. F., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage*, *31*, 1257–1267.

Crutcher, M. D., & Alexander, G. E. (1990). Neural representations of the target (goal) of visually guided arm movements in three motor areas of the monkey. *J. Neurophysiol.*, *64*, 151–163.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *Eur. J. Neurosci.*, *15*, 399–402.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *J. Neurophysiol.*, *73*, 2608–2611.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur. J. Neurosci.*, *17*, 1703–1714.

Ferrari, P. F., Rozzi, S., & Fogassi, L. (2005). Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *J. Cogn. Neurosci.*, *17*, 212–226.

Filimon, F., Nelson, J. D., Hagler, D. J., & Sereno, M. I. (2007). Human cortical representations for reaching: Mirror neurons for execution, observation, and imagery. *NeuroImage*, *37*, 1315–1328.

Fogassi, L., Bonini, L., Simone, L., Ugolotti, F., Ruggeri, E., Rozzi, S., Chersi, F., Rizzolatti, G., & Ferrari, G. (2007). Time course of neuronal activity reflecting the final goal of observed and executed action sequences in monkey parietal and premotor cortex. *Soc. Neurosci. [Abstracts]*, 636.4.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, *302*, 662–667.

Frith, C. D., Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron*, *6*, 503–510.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci.*, *8*, 396–403.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007a). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, *35*, 1674–1684.

Gazzola, V., van der Worp, H., Mulder, T., Wicker, B., Rizzolatti, G., & Keysers, C. (2007b). Aplasics born without hands mirror the goal of hand actions with their feet. *Curr. Biol.*, *17*, 1235–1240.

Gregoriou, G. G., Borra, E., Matelli, M., & Luppino, G. (2006). Architectonic organization of the inferior parietal convexity of the macaque monkey. *J. Comp. Neurol.*, *496*, 422–451.

de Hamilton, A. F., & Grafton, S. T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cereb. Cortex*, *18*, 1160–1168.

Hoshi, E., & Tanji, J. (2000). Integration of target and body-part information in the premotor cortex when planning action. *Nature*, *408*, 466–470.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J., & Rizzolatti, G. (2005). Grasping the intentions of others with one's owns mirror neuron system. *PLoS Biol.*, *3*, 529–535.

Jastorff, J., Rizzolatti, G., & Orban, G. (2007). Somatotopy vs actinotopy in human parietal and premotor cortex. *Soc. Neurosci. [Abstracts]*, 127.1.

Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. *Trends Neurosci.*, *18*, 314–320.

Kakei, S., Hoffman, D. S., & Strick, P. L. (1999). Muscle and movement representations in the primary motor cortex. *Science*, *285*, 2136–2139.

Kakei, S., Hoffman, D. S., & Strick, P. L. (2001). Direction of action is represented in the ventral premotor cortex. *Nat. Neurosci.*, *4*, 1020–1025.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*, 846–848.

Leiguarda, R. C., & Marsden, C. D. (2000). Limb apraxias: Higher-order disorders of sensorimotor integration. *Brain*, *123*, 860–879.

Maestripieri, D. (1996). Gestural communication and its cognitive implications in pigtail macaques (Macaca nemestrina). *Behaviour*, *133*, 997–1022.

Matelli, M., Camarda, R., Glickstein, M., & Rizzolatti, G. (1986). Afferent and efferent projections of the inferior area 6 in the macaque monkey. *J. Comp. Neurol.*, 251, 281–298.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex*, *18*, 262–271.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *J. Neurophysiol.*, *78*, 2226–2230.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., & Sakata, H. (2000). Selectivity for the shape, size and orientation of objects in the hand-manipulation-related neurons in the anterior intraparietal (AIP) area of the macaque. *J. Neurophysiol.*, *83*, 2580–2601.

Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2005). Observing others: Multiple action representation in the frontal lobe. *Science*, *310*, 332–336.

Oberman, L. M., Hubbard, E. H., McCleery, J. P., Altschuler, E., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cogn. Brain Res.*, *24*, 190–198.

Orban, G., Peeters, R., Nelissen, K., Buccino, G., Vanduffel, W., Rizzolatti, G. (2006). The use of tools, a unique human feature represented in the left parietal cortex. *Soc. Neurosci.* [*Abstracts*], 144.2.

Pandya, D. N., & Seltzer, B. (1982). Intrinsic connections and architectonics of posterior parietal cortex in the rhesus monkey. *J. Comp. Neurol.*, *204*, 196–210.

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., Chitty, A. K., Hietanen, J. K., & Ortega, J. E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *J. Exp. Biol.*, *146*, 87–113.

Raos, V., Umilta, M. A., Fogassi, L., & Gallese, V. (2006). Functional properties of grasping-related neurons in the ventral premotor area F5 of the macaque monkey. *J. Neurophysiol.*, *95*, 709–729.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.*, *21*, 188–194.

Rizzolatti, G., & Craighero, L. (2004). The mirror neuron system. *Annu. Rev. Neurosci.*, *27*, 169–192.

Rizzolatti, G., Camarda, R., Fogassi, M., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements. *Exp. Brain Res.*, *71*, 491–507.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.*, *3*, 131–141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 539–552). Cambridge, MA: MIT Press.

Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: New concepts. *Electroencephal. Clin. Neurophysiol.*, *106*, 283–296.

Rozzi, S., Calzavara, R., Belmalih, A., Borra, E., Gregoriou, G. G., Matelli, M., & Luppino, G. (2006). Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cereb. Cortex*, *16*, 1389–1417.

Rozzi, S., Ferrari, P. F., Bonini, L., Rizzolatti, G., & Fogassi, L. (2008). Functional organization of inferior parietal lobule convexity in the macaque monkey: Electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. *Eur. J. Neurosci.*, *28*, 1569–1588.

Sakreida, K., Schubotz, R. I., Wolfensteller, U., & von Cramon, D. Y. (2005). Motion class dependency in observers' motor areas revealed by functional magnetic resonance imaging. *J. Neurosci.*, *25*, 1335–1342.

Saygin, A. P., Wilson, S. M., Hagler, D. J. Jr., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *J. Neurosci.*, *24*, 6181–6188.

Schmitz, C., Martineau, J., Barthélemy, C., & Assaiante, C. (2003). Motor control and children with autism: Deficit of anticipatory function? *Neurosci. Lett.*, *348*, 17–20.

Shmuelof, L., & Zohary, E. (2006). A mirror representation of others' actions in the human anterior parietal cortex. *J. Neurosci.*, *26*, 9736–9742.

Theoret, H., Halligan, E., Kobayashi, M., Fregni, F., Tager-Flusberg, H., & Pascual-Leone, A. (2005). Impaired motor facilitation during action observation in individuals with autism spectrum disorder. *Curr. Biol.*, *15*, 84–85.

Tkach, D., Reimer, J., & Hatsopoulos, N. G. (2007). Congruent activity during action and action observation in motor cortex. *J. Neurosci.*, *27*, 13241–13250.

Ulloa, E. R., & Pineda, J. A. (2007). Recognition of point-light biological motion: Mu rhythms and mirror neuron activity. *Behav. Brain Res.*, *183*, 188–194.

Umiltà, M. A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., et al. (2008). How pliers become fingers

in the monkey motor system. *Proc. Natl. Acad. Sci. USA*, *105*, 2209–2213.

Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). "I know what you are doing": A neurophysiological study. *Neuron*, *32*, 91–101.

Van Hoof, J. A. R. A. M. (1962). Facial expressions in higher primates. *Symp. Zool. Soc. Lond.*, *8*, 97–125.

Van Hoof, J. A. R. A. M. (1967). The facial displays of the catarrhine monkeys and apes. In D. Morris (Ed.), *Primate ethology* (pp. 7–68). London: Weidenfield & Nicolson.

Wheaton, L. A., Carpenter, M., Mizelle, J. C., & Forrester, L. (2008). Preparatory band specific premotor cortical activity differentiates upper and lower extremity movement. *Exp. Brain. Res.*, *184*, 121–126.

Wheaton, L. A., & Hallett, M. (2007). Ideomotor apraxia: A review. *J. Neurol. Sci.*, *260*, 1–10.

Yokochi, H., Tanaka, M., Kumashiro, M., & Iriki, A. (2003). Inferior parietal somatosensory neurons coding face-hand coordination in Japanese macaques. *Somatosensory Mot. Res.*, *20*, 115–125.

44 Relative Hierarchies and the Representation of Action

SCOTT T. GRAFTON, L. AZIZ-ZADEH, AND R. B. IVRY

ABSTRACT Hierarchy is a central concept for understanding how complex goal-oriented behaviors are organized. In this chapter we present recent functional and behavioral evidence that supports the existence of a control hierarchy in the human brain for organizing complex motor. It is proposed that the functional hierarchy is not based on strict anatomical connectivity within the motor system. Instead, there are multiple motor planning circuits, each of which can serve a supraordinate role, and this role can be readily interchanged to achieve a much wider range of task outcomes. This can be observed at the level of hand-object interactions, bimanual control, and the integration of semantics into action planning.

Historical perspective: The hierarchy of serial behavior

Within the field of motor control, the problem of how people accomplish complicated tasks has historically been intertwined with the concept of what constitutes a motor program. This in turn depends on solving the problem of serial order, first articulated by Lashley (1951). He sought to understand how the nervous system organized sequential motor elements to achieve a desired motor goal. Bernstein (1996) elaborated on the serial order problem by emphasizing that the control system was flexible, designed to produce actions that were constrained by task demands rather than fixed action patterns. This shift in emphasis brought to the forefront the concept of a goal. We move to accomplish goals.

These early theories were critical in minimizing the role of the simple chaining of reflexes, proposing instead the existence of a motor plan as an alternative. But they also introduced fundamental questions that continue to challenge the field: What is a motor plan? Is it composed of discrete representational elements? Can associative mechanisms that underlie reflex chains be used to solve more complex problems, including those associated with task planning? Do all plans require goals? How does a task get organized?

One possible mechanism for planning goal-directed sequences of actions is based on hierarchical control. The argument for hierarchical control was elaborated within a cognitive science framework by Keele and colleagues (1990) during the 1980s. Using a set of model tasks such as handwriting and the serial reaction time task, researchers sought to describe the structure of control hierarchies by identifying consistent patterns of variation in the time required to initiate successive components of an action, as well as through studies of motor transfer. These studies showed that many aspects of control reflected constraints that were related to the abstract nature of action representations, separable from the musculoskeletal system. A fundamental distinction derived from this perspective can be made between abstract plans and their implementation, the basic components of a hierarchy.

Another form of hierarchical control within the implementation process itself became evident through studies of naturalistic grasping. Kinematic analysis showed that the transport phase of an arm movement has an exquisite interdependency with processes involved in shaping the hand to grasp an object (Jeannerod, 1984, 1986). The velocity of the transport phase is subordinate to the grasp requirements, with timing that is tightly coupled to maximal hand aperture. At a more abstract level, prior experience and task goals can also influence grasping (Rosenbaum, Meulenbroek, & Vaughan, 2001; Rosenbaum, Vaughan, Barnes, & Jorgensen, 1992). For example, the way in which an object is grasped is constrained by task demands. Given a fixed starting position, the adopted grasping posture will depend on how the tool is to be moved (defined by the center of mass) and used (defined by the tool's functional properties), as well as the comfort of the end-state posture. In this case the selected grip is subordinate to the desired goal for using the tool as well as biomechanical constraints.

Computational models of motor planning have also exploited hierarchical features in action representation. In a model of hierarchical behavior proposed by Cooper and Shallice (2006), a logical tree structure of discrete behaviors is developed to organize an action sequence. To make a cup

SCOTT. T. GRAFTON UCSB Brain Imaging Center, Department of Psychology, University of California Santa Barbara, Santa Barbara, California

L. AZIZ-ZADEH Brain and Creativity Institute and The Department of Occupational Therapy, University of Southern California, Los Angeles, California

R. B. IVRY Department of Psychology, University of California Berkeley, Berkeley, California

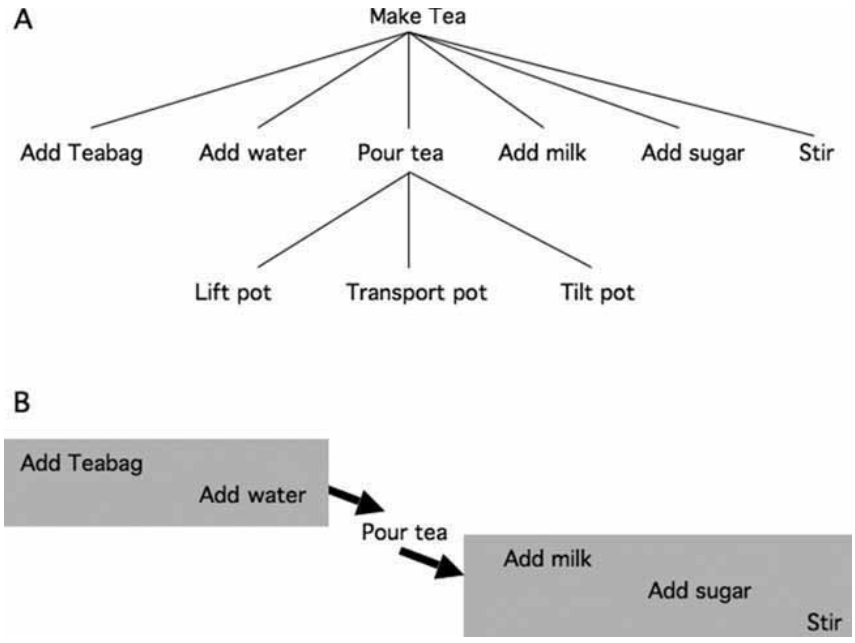


FIGURE 44.1 Task hierarchy for making a cup of tea. (*A*) An explicit whole-part schema for performing a task with three levels of complexity. (*B*) A schema of the same task after the different components have been compiled into sequential units.

of coffee, the act of adding sugar is distinct from the act of adding milk, and each must be scheduled after the coffee has been brewed. This scheduling occurs within a large multilayered, interactive network, with the top level of the hierarchy providing constraint in terms of its specification of task goal (see figure 44.1*A*).

In such models, the notion of a hierarchy is explicit, with the layered representation defined as a task schema. In an alternative approach, action planning could be goal-independent, with the hierarchy arising as an emergent property of processes that arises from sequential transitions between different components. For example, Botvinick (2008) has shown that a simple recursive network based on an action layer, a perception layer, and an intermediate layer can learn fairly complex motor actions without the need for top-down task structuring with respect to a goal. Furthermore, sequencing and the formation of motor programs can lead to compilation of complex acts into a smaller set of tasks, as shown in figure 44.1*B*. The evaluation of these computational models has primarily relied on behavioral studies that involve dependent variables such as variation in planning time and errors of substitution.

Methods of cognitive neuroscience can provide additional means for addressing these issues. The present chapter seeks to incorporate these other forms of evidence to reconsider how the notion of hierarchy and action goals may be conceptualized to aid our understanding of how movement is achieved. To assess these questions, we review a wide range of actions, spanning tool use, bimanual coordination, and, finally, how language influences motor planning and control.

Anatomic versus representational hierarchy

The concept of an action hierarchy has not been limited to the psychological and computational realms; neuroscientists have also been highly influenced by hierarchical notions in theorizing about the organization of the nervous system. Efforts to map different levels of motor planning into distinct neural substrates were motivated in large part by the belief that there existed a strict anatomical hierarchy, at least within lower levels of the nervous system. As one ascends from muscle activity to peripheral nerves and then into spinal cord and ultimately to motor cortex, there is increasing abstraction in the type of information represented (d'Avella & Bizzi, 2005; Giszter, Mussa-Ivaldi, & Bizzi, 1993). It is only natural to assume that there is a continuation of this control hierarchy into premotor and ultimately prefrontal and parietal areas. An example is the sensorimotor hierarchy first proposed by Fuster (Fuster, 1995) and recently implemented by Botvinick (2007) (figure 44.2*A*). In this model, only the primary motor cortex influences the environment. As a task becomes more complex, increasing reliance is placed on premotor heteromodal sensory circuits and ultimately prefrontal-polymodal sensory circuits. In an extreme version of this model, there is a linear gradient between task complexity and posterior to anterior prefrontal cortex (Badre & D'Esposito, 2007; Botvinick, 2007, 2008).

Early brain imaging studies of action planning were often interpreted as being consistent with an anatomical hierarchical framework (Roland, Larsen, Lassen, & Skinhoj, 1980;

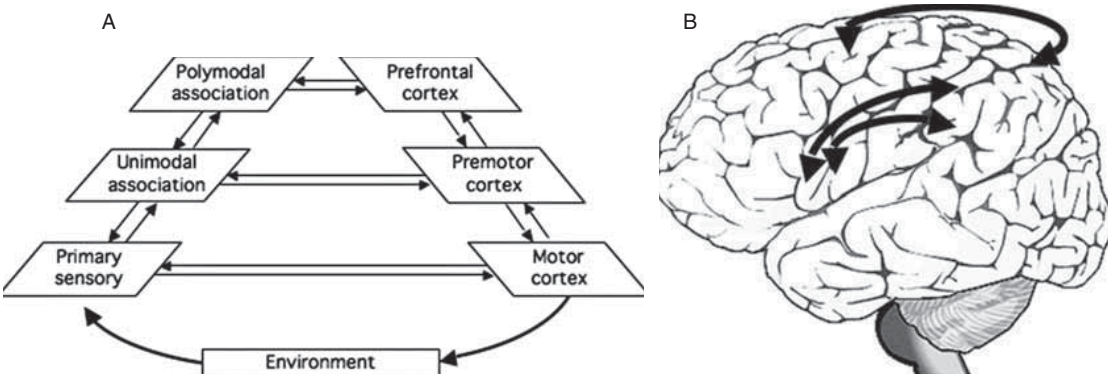


FIGURE 44.2 Examples of anatomic networks. (*A*) In this model, based on Fuster (1995), sensorimotor loops represent information of increasing abstraction or complexity. Only the motor cortex controls interactions with the environment, and there is strict segregation between unimodal and polymodal sensory areas. (*B*) In this multiple stream model, there are at least two parietal-prefrontal-premotor streams engaged for goal-oriented behavior. A dorsal-dorsal stream (upper arrow) is used for learning arbitrary sensorimotor transformations and reaching. A ventral-dorsal stream (middle arrow) is used for object centered actions. A third stream, positioned between inferior parietal lobule and inferior frontal gyrus (lower arrow), has been hypothesized for representing complex actions and tool use. In this model, these circuits operate in tandem, with no fixed hierarchical arrangement.

Roland, Skinhøj, Lassen, & Larsen, 1980). One of these studies used positron emission tomography to measure blood flow and compared activation patterns during real versus imagined movement. Whereas the supplementary motor area (SMA) was active during both real and imagined movement, motor cortex was only weakly activated during imagined movements. This dissociation was interpreted as showing that the SMA provided a more abstract representation of the action plan, one that provided the plan of the action, and motor cortex activation was primarily limited to the actual implementation of that plan. From this view, the SMA has sometimes been referred to as a "supramotor" area.

However, four arguments suggest that caution is required in attempting to identify a direct correspondence between a well-defined anatomical hierarchy and a functional hierarchy. First, many of the descending pathways to the spinal cord originate outside the primary motor cortex, including rich projects from premotor and parietal cortex, as well as the extrapyramidal brain stem pathways (Dum & Strick, 1991, 1996). The diversity of these cortical and subcortical projections underscores the ability of these areas to directly influence movement. In addition to their direct influences on motor commands, these areas likely play a role in establishing the context of an action and the coordination of movement commands with current information about the state of the actor.

A second argument against the presence of a strict anatomical hierarchy within motor regions of the cortex is that most premotor areas have direct inputs onto motor cortex, and no premotor area appears to have a dominant role over another (Dum & Strick, 2005). It has become clear that there are multiple body representations within premotor cortex, and the anatomy fails to indicate some sort of hierarchical structure across these subregions.

The third argument is based on recent evidence that even the lowest levels of this presumed cortical hierarchy are capable of organizing extremely complex serial behavior (Lu & Ashe, 2005; Matsuzaka, Picard, & Strick, 2007). Recent recordings from the primary motor cortex in nonhuman primates demonstrate sequence-specific responses that are tied to the action rather than particular muscles. Similar evidence for learning-dependent changes within motor cortex for complex serial actions has been observed in humans (Grafton, Hazeltine, & Ivry, 1998; Grafton, Salidis, & Willingham, 2001; Karni et al., 1998).

Finally, models that emphasize a strict anatomical hierarchy for motor planning run the risk of requiring a command and control structure with a "decider" at the top; the problem of the homunculus resurfaces in such models. This type of architecture seems difficult to reconcile with the effortless nature with which we perform many of our everyday actions. These are planned unconsciously and adjusted on-line at an extremely rapid rate (Desmurget & Grafton, 2000).

An alternative conceptualization of functional anatomy is motivated by the existence of multiple interactive loops across prefrontal-parietal cortex. For example, the concept of two visual streams for object identification versus action pragmatics has been extended, as is shown graphically in figure 44.2*B* (Goodale, Milner, Jakobson, & Carey, 1991). There is now extensive anatomical and functional evidence to support at least two and possibly three processing streams within the classic "dorsal" stream related to object-centered action, tool use, and reaching (Johnson & Grafton, 2003; Rizzolatti & Luppino, 2001; Rizzolatti & Matelli, 2003). In addition, there is little anatomical evidence to segregate polymodal sensory from unimodal association cortex as originally proposed by Fuster. Within each parietal-premotor-prefrontal pathway, all forms of sensory information are integrated.

The preceding arguments suggest that insight into the hierarchical nature of action planning and goal representation will not be defined by the existence of an anatomical hierarchy. Instead, an anatomical organization with multiple parallel parietal-prefrontal and premotor pathways supports a multitude of relative hierarchies that can be flexibly recruited as a function of task demands, experience, and context. In this framework, there are dissociable functional anatomic substrates, but these are not constrained by a fixed hierarchy. This shifts the focus of inquiry to understanding representational hierarchies that are highly flexible and goal based.

Goal representation and the on-line control of grasp

Grasping studies traditionally focus on the interplay between grip formation and limb transport to understand the representational organization of these two task components (Haggard & Wing, 1997; Jeannerod, 1997; Jeannerod, Arbib, Rizzolatti, & Sakata, 1995). For grasping, the object itself defines the task goal. The problem then becomes one of sensorimotor transformation, in which object features are decoded to generate hand configurations that are optimally shaped to match the object geometry. Object knowledge involves both physical properties (texture, mass, center of gravity) and utility (how the parts of an object such as a handle and action surfaces are used to accomplish particular goals). Through experience and cumulative knowledge, a library of possible hand-object affordances and utility are constructed.

Neuropsychological and neuroimaging studies have provided evidence of two primary pathways in posterior cortex: the classic "how" and "what" visual streams (Culham et al., 2003; Goodale et al., 1994). Processing distinctions within these pathways can be viewed as supporting pragmatic and conceptual representations for action. Within the dorsal "how" pathway, there is clear evidence that the anterior intraparietal sulcus (aIPS) in humans (area AIP in nonhuman primates) is critical for sensorimotor transformations that relate the visual and/or haptic features of an object to a desired hand shape, with limb transport, body stabilization, and eye movements playing subordinate roles.

This framing in terms of a sensorimotor transformation provides a starting point to understand a functional hierarchy, but it is missing a critical piece: how the goal is related to the sensorimotor transformations. We grasp objects to achieve goals (e.g., pick up a nut) or to solve problems (e.g., use a tool to open the nut). How does an area such as aIPS integrate the low-level details required to control grip aperture with information about an object that includes high-level features that may be defined functionally and in a way that varies with context? This shifts the problem from one of sensorimotor transformation to one of sensorimotor integration (Wolpert, Ghahramani, & Jordan, 1995; Wolpert, Goodbody, & Husain, 1998).

There are at least two solutions to this problem. One is that areas such as aIPS are akin to low-level visual areas and pass information off to higher cortical areas. This is a classic functional-anatomical hierarchy of ascending representational complexity. For example, the goal level of the action might be represented in ventral premotor cortex, an area that is richly connected to aIPS. In this framework, premotor areas would make the ultimate control decisions. An alternative view is that of an inverted or flexible hierarchy. This emphasizes that information about the task goal can have a direct influence on processing within areas such as aIPS. In this scheme, computations within aIPS use this goal information to constrain sensorimotor integration needed to related motor commands with object information and an internal representation of the body in relationship to the object.

A series of transcranial magnetic stimulation (TMS) studies targeting aIPS during grasping suggest that the latter perspective is more appropriate. In this work, subjects were required to reach and grasp a small $1 \times 1 \times 5$ cm rectangular wooden block located on a computer-controlled torque motor (Tunik, Frey, & Grafton, 2005). The orientation of the block could be changed in less than 30 ms. The subject was required to start with the right hand on a button and, when ready, use the index finger and thumb to grasp the object, aligning these fingers on the vertical axis. To assess on-line updating, the initiation of the grasping movement would trigger the motor to spin the object 90 or 180 degrees. In this manner, the object's orientation always changed on every trial. However, the planned grasping action would have to be updated when the object rotated 90 degrees, requiring a larger aperture to grasp the 5-cm width. In contrast, when the object rotated 180 degrees, no adjustment was required, since the object's vertical axis remained at 1 cm diameter.

Single pulses of TMS were delivered to aIPS, timed to the start of the hand movement. The TMS pulses disrupted the subjects' ability to modify their grip aperture on trials in which the grasp had to be updated but had no effect on trials in which the original grip could be used (figure 44.3). In control conditions in which the TMS pulses were applied to other brain regions (e.g., caudal or mid intraparietal cortex), or 400 ms after movement onset, no behavioral effects were observed.

If reaching and grasping are anatomically dissociable, TMS of aIPS should affect only the control of grip aperture. To test this hypothesis, the same subjects were tested in a critical second experiment. Instead of always grasping the block along the vertical axis, they were now told to always grip the narrow (1-cm) axis. With these instructions, the task goal required that they update the orientation of the forearm and wrist, but the grip aperture remained fixed. In this con-

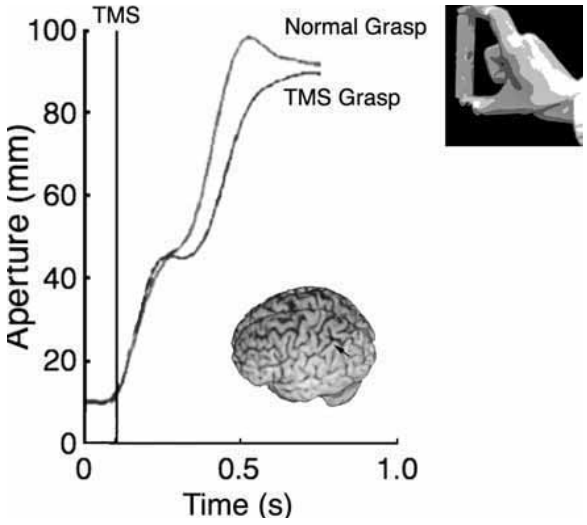


FIGURE 44.3 Effect of a single pulse of TMS during object grasping. The figure plots finger aperture measured when subjects grasp an object that has increased in size, as shown in the insert of the hand. When TMS is applied to the anterior intraparietal sulcus (arrow on brain, insert) at movement onset, there is a delay in the formation of the required grip aperture (lower curve in plot) compared to the no-TMS, control condition (upper curve in plot). (Adapted from Tunik et al., 2005.)

dition, when TMS pulses were applied to aIPS at movement onset, the participants were unable to rotate the arm appropriately to match the new orientation of the object.

Taken together, the results argue against the hypothesis that the TMS is interfering with the adjustment of a particular set of muscles or an elemental process such as grip aperture. A more parsimonious explanation is that this region of the parietal cortex is involved in using information concerning the task constraints to generate a desired hand-object interaction. As such, the TMS appears to disrupt the representation of the goal itself rather than some "downstream" operation that controls some component of that goal.

Note that in the two studies reviewed above, the disruptive effects of TMS were observed on trials in which an action plan had to be updated rapidly. Perhaps aIPS is important for this updating process and not needed when the action has already been planned. To address this, shutter goggles were used to control when visual information about the object was available to the subjects (Rice, Tunik, & Grafton, 2006). The goggles provided a brief 200-ms view of the object but were closed just before reach onset. Thus the subjects were required to reach and grasp the object without vision of the hand or of the object. Critically, TMS to aIPS interfered with grasp kinematics when the pulses were delivered at movement onset but not when it was delivered during the viewing period.

This result argues strongly against the hypothesis that aIPS is essential for planning the requisite sensorimotor transformation solely on the basis of visual information. If this were the case, disruption should also have been observed when the pulses were applied before movement onset. Instead, aIPS appears to be essential for the on-line integration of visual and proprioceptive information with efference copy signals in order to meet the task goal.

The TMS findings are consistent with the hypothesis that the task goal is embedded in the controllers, even for relatively simple actions such as grasping. Moreover, the constraints associated with the task goal are closely intertwined with processes involved in the sensorimotor integration needed for rapid adjustments of an ongoing movement. At least with respect to the role of aIPS in grasping, there is no evidence for an anatomical dissociation between the representation of a goal and the implementation of movements required to achieve that goal.

Goal representation in bimanual coordination

Studies of bimanual movements have provided another framework for examining how the selection and control of movement are constrained by action goals. Much of this work has involved rhythmic movements, evaluating changes in pattern stability when performers are asked to adopt a range of phase relationships between the two hands (see Schöner & Kelso, 1988). A large body of evidence demonstrates that certain patterns are more stable than others; people are much more adept in adopting antiphase and in-phase patterns of motion than in adopting patterns in which the two hands must adopt more complex phasing patterns. These constraints have been formally described by models in which the limbs are conceptualized as coupled oscillators. An alternative, process-oriented perspective focuses on the manner in which the task goals are represented. In the event-timing model of Spencer, Semjen, Yang, and Ivry (2007), the task is represented as a series of salient temporal events such as the point of contact during finger tapping or maximum flexion or extension during movements performed without such haptic cues. In this model, stability is constrained by the complexity of the temporal representation. Thus, antiphase and in-phase patterns are more stable because the representation of the temporal goals for these patterns is simpler than for more complex phase relationships (Semjen & Ivry, 2001). Moreover, the phase transitions observed from antiphase to in-phase movement when movement frequency increases arise because the latter entails a simpler temporal representation (Spencer et al., 2007).

A similar perspective has been offered to account for bimanual interactions observed in the spatial domain (Ivry, Diedrichsen, Spencer, Hazeltine, & Semjen, 2004). Consider a task in which a person must simultaneously draw two three-sided squares (figure 44.4*A*). Performance is fluid when the patterns are symmetric (e.g., U and U). In contrast, when the patterns are orthogonal (e.g., U and C), severe limitations

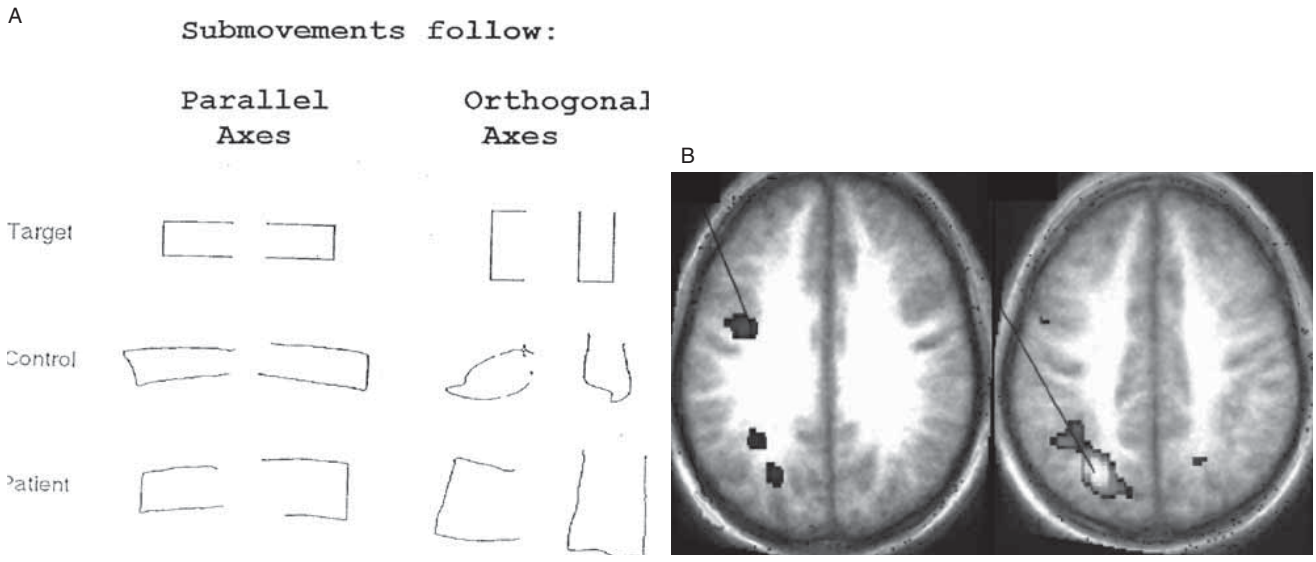


FIGURE 44.4 (*A*) In this task, the participant must simultaneously draw the shape on the left with the left hand and the shape on the right with the right hand. Representative trajectories produced by a control participant and a callosotomy patient are shown. (*B*) Symbolically cued actions produce stronger activation across the left intraparietal sulcus/superior parietal lobule and left premotor cortex than do directly cued actions. (Adapted from Diedrichsen et al., 2006.)

are observed; the time to initiate each segment increases dramatically, and spatial distortions are observed such that the trajectories for the two hands become assimilated (Albert & Ivry, 2009; Franz, Eliasson, Ivry, & Gazzaniga, 1996). These effects can also be observed with simpler patterns; for example, reaction times are much slower when two linear movements are symmetric than when they follow orthogonal trajectories or are of different amplitudes (Heuer, Kleinsorge, Spijkers, & Steglich, 2001). However, these costs are essentially abolished when stimuli appear at the endpoint location, serving as direct cues for the required movements (Diedrichsen, Hazeltine, Kennerley, & Ivry, 2001).

The fact that the constraints are highly dependent on the manner in which the actions are cued indicates that the limitations here are not related to processes that are typically associated with motor programming and execution. Rather, they arise at a more abstract level, one associated with the goal of the action. This may be related to the sensory consequences of the movements (Franz, Zelaznik, Swinnen, & Walter, 2001; Mechsner, Kerzel, Knoblich, & Prinz, 2001) or to the manner in which the movements themselves are conceptualized (e.g., as movements to produce trajectories or movements to locations; see Ivry et al., 2004).

The neural locus of bimanual coordination has been the subject of considerable study. Within a traditional hierarchical perspective, the debate has centered on whether certain neural regions are specialized for coordinating the gestures of the two hands. Much of this work has focused on the supplementary motor area, motivated by anatomical, lesion, and neuroimaging evidence involving sequential or rhythmic movements (e.g., Brinkman, 1984; Sadato, Yonekura, Waki, Yamada, & Ishii, 1997). Diedrichsen and colleagues (2006) looked at simpler movements, comparing conditions in which reaching movements were either directly cued (e.g., targets specified by the locations of the stimuli) or symbolically cued (e.g., locations specified by letters indicating target locations). The SMA showed no difference between conditions requiring unimanual or bimanual movements. However, a large parietal region extending along the intraparietal sulcus as well as premotor cortex showed greater activation in the symbolic condition than in the direct condition (figure 44.4*B*). Interestingly, this activation was much stronger in the left hemisphere than in the right hemisphere, and the magnitude of the activation was greater for bimanual movements that were incongruent (e.g., orthogonal directions) than when they were congruent (e.g., parallel conditions).

In terms of the focus of this chapter, these findings speak to three issues. First, contrary to predictions derived from a traditional anatomical-inspired framework, there was no region, including SMA, that appeared to be specifically sensitive to the contrast of unimanual and bimanual movements. Thus, at least for reaching, the evidence fails to support the hypothesis that there exists a neural region that is specialized for bimanual coordination. Second, manipulation of the task goals did not engage new neural regions but rather led to modulation of the magnitude of neural activity. That is, a similar network was recruited for reaching movements in response to direct and symbolic cues, the activation in these areas being greater in the latter conditions. Thus, similar to our review of reaching and grasping, the representation of the task goal and control

operations required to achieve that goal appear to be intimately intertwined.

Third, the symbolic condition did reveal a strong hemispheric asymmetry. Activation was strongly lateralized to the left hemisphere for action goals that required a translation between stimuli and their associated responses. At a functional level, this latter finding suggests some form of hierarchical organization between the two hemispheres, with the left hemisphere playing a dominant role when there is some degree of abstract representation of the action goals.

Action understanding

The hierarchical representation of action has also been addressed in studies of action comprehension. As with studies of perceptual organization, we can ask how people attend to actions and organize their percepts. Do we focus on the goals of an action, perhaps at a cost of becoming aware of the component parts (e.g., the specifics of the gestures)? Or must we process the components in detail to arrive at an understanding of the actor's intentions?

Conventional subtractive methodologies are problematic in functional magnetic resonance imaging (fMRI) studies of hierarchical organization. It is difficult to manipulate how people perform action comprehension tasks, and doing so is likely to affect a host of factors, such as task complexity and the demands on working memory. An alternative approach is to exploit the repetition suppression (RS) phenomenon, a tool that has proven useful across a range of task domains in perception (Grill-Spector & Malach, 2001; Kourtzi & Kanwisher, 2000). RS is also referred to as *fMRI adaptation*. When two successive stimuli are "similar," the BOLD response to the second stimulus is reduced (i.e., a form of adaptation or habituation) compared to conditions in which the second stimulus is novel (i.e., release from adaptation).

The key here is that RS provides a signature of what "similar" means across the brain. For example, in face perception, some brain regions may show an RS effect only when the gender of the face remains the same on successive trials; the same or other regions may show RS only when the same face is presented. As such, RS provides a probe to identify hemodynamic changes within a class of stimuli or a level of a hierarchy rather than between classes. In this way, different levels of representation for the same stimulus can be analyzed independently.

To identify topologies in the human brain corresponding to a motor hierarchy, Hamilton and Grafton (2006) had participants view short video clips of a right arm reaching and grasping either of two objects and then had them transport the selected object to a midline position. Two variables were manipulated. First, one or the other object was grasped. Second, the position of the two objects with respect to the midline was varied. In this way, repetition could be defined in terms of the trajectory or the object that was grasped. A strong RS effect was observed in the left anterior intraparietal sulcus (aIPS) when the same object was grasped. Importantly, the aIPS was not sensitive to trajectory repetition. Instead, RS effects based on trajectory were limited to the left lateral occipital sulcus and right superior precentral sulcus. This double dissociation provides strong evidence in favor of some type of hierarchy during action observation, with one level sensitive to kinematic-like features of the action (i.e., trajectory) and another sensitive to the goal of the action, defined here in terms of which object is grasped.

In a follow-up experiment, subjects observed movies of an actor reaching out to grasp a wine bottle or a dumbbell; the latter object was placed on end such that its primary axis matched that of the wine bottle (Hamilton & Grafton, 2007). As in the first experiment, the object to be grasped either remained the same or changed from one trial to the next. Independent of this, the form of the grasp was manipulated by having the actors use either a power grip, in which the object was grasped along its thick part, or with a pincer grip, in which the object was grasped along its thin part. The RS effect related to which object was grasped was again localized to the left aIPS (extending into the adjacent IPL) and to a lesser degree within the right aIPS. In contrast, an RS effect based on how the object was grasped was observed in the SMA, middle IPS, inferior and middle occipital regions, and clusters within middle and inferior frontal gyri. These results again suggest a hierarchical organization with some regions sensitive to the form of the grasp and others sensitive to the goal of the action.

In the preceding studies, the goal of the action was defined as the object to be grasped, with the form of the action considered subordinate. However, in most actions, grasping is really the means to an end: We grasp objects such as tools to accomplish some behavioral goal that requires manipulation of that object (Frey, 2008). Thus, the RS experiments described above are not adequate for assessing intentionality in terms of why the actor has selected one object over the other. This issue was addressed in an RS experiment that manipulated the outcome of the actions (Hamilton & Grafton, 2008). For example, a series of video clips were generated in which the actor pushed or pulled a sliding cover that was attached to the top of a wooden box. Depending on the starting position of the lid, the result of this movement either opened or closed the box. A similar manipulation of kinematics and outcome was used with a range of familiar tasks such as turning a stove on or off, tying or untying a string, and drawing or erasing with a pencil.

An RS effect for outcomes was found in the bilateral IPL and the IFG. These effects were not driven by a single action or outcome but were generalized across a wide variety of actions. As in the other RS experiments, these areas were not sensitive to the means through which the actions were

achieved. Rather, repetition of the kinematics (e.g., push or pull) led to a reduced BOLD response in the left middle intraparietal sulcus, left lateral occipital cortex, and left superior temporal sulcus.

Taken together, these three experiments support a model of representational hierarchy that distinguishes action means, kinematics, object-centered behavior, and ultimately, action consequences. The decoding of object-centered action appears to be strongly left lateralized, whereas the decoding of more complex action intentions arising as a consequence of the action engaged bilateral frontal-parietal circuits. The bilateral recruitment that is observed in this latter condition is quite different from the relative hierarchies described in the other sections of this chapter. One explanation focuses on perceptual factors. Complex intentions might require more global perceptual analysis (Ivry & Robertson, 1998). An alternative explanation is that the right hemisphere plays a central role in representing more complex action goals. Most studies of action understanding or production focus on simple object-centered actions rather than complex goals and do not address this hypothesis. This explanation is supported by patient studies. As Hartmann, Goldenberg, Daumüller, and Hermsdörfer (2005, p. 625) recently emphasized, "It takes the whole brain to make a cup of coffee."

Action semantics

How might language semantics fit into the representational hierarchy of motor control? It seems plausible that a word such as *hammering* could summon the actions associated with this concept. Thus, when one hears the word, an entire action plan would be activated, one composed of various subcomponents: retrieving the required tools, grasping the hammer with one hand and the nail with the other, striking the nail by pounding the hammer. The hypothesis of an interaction between semantic processing and action planning is supported by evidence from various methodologies. For example, adjectives related to object properties have been found to influence movement execution (Gentilucci, Benuzzi, Bertolani, Daprati, & Gangitano, 2000; Gentilucci & Gangitano, 1998; Glover & Dixon, 2002). A subject's initial grasp kinematics is influenced by seeing the word *large* or *small* printed over the target object. Similarly, initial reach kinematics to an object are altered if the word *far* or *near* is printed adjacent to the object. These findings indicate that semantic processing, even when not explicitly related to the motor task, influences motor planning. As such, they reveal how language provides another representational system through which motor plans are organized and influenced.

The form of these interactions has been the focus of numerous recent investigations. One hypothesis is that language and motor systems constitute two parallel systems (figure 44.5). Task requirements determine which system is

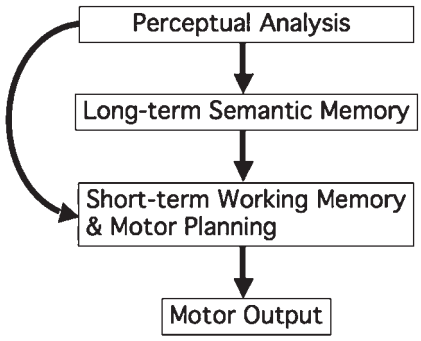


FIGURE 44.5 Two routes, one mediated by semantics, for building up a motor representation. (Adapted from Tessari & Rumiati, 2004.)

used. For example, consider a task in which a person is asked to imitate gestures. If the gestures are meaningless, then it is thought that imitation must occur via a direct visuomotor route. If the gestures are meaningful, however, then imitation could be achieved either by this direct visuomotor route or by accessing long-term semantic memory (Rumiati & Tessari, 2002; Rumiati et al., 2005). Behavioral studies of imitation of meaningful and meaningless gestures (Tessari & Rumiati, 2004) support the theory that actions can be organized by these two systems.

Is processing in these pathways independent, or do the systems share some common neural substrates? A visuomotor route would seem to involve motor-related areas. To what degree does a semantic route use (some of) the same motor-related brain regions? Does hearing the word *hammering* directly activate motor-related brain areas? One way to explore this question is to return to the study of action comprehension. If an action-related semantic area is independent of motor-related areas, then comprehension should remain possible if the motor regions are damaged, at least for meaningful actions. In this view, words related to actions are processed by nonmotor, language-related areas; their effect on motor performance is indirect, perhaps occurring via spreading activation to motor regions. Alternatively, if action semantics is intimately linked with motor-based representations, then lesions of the motor regions should disrupt comprehension. That is, in this view, semantic knowledge cannot be separated from the systems that are involved in producing the actions themselves, a form of embodied cognition (e.g., Gallese & Lakoff, 2005). As such, lesions to these areas will affect both action production and action comprehension.

This question has been asked in a number of neuropsychological studies. Some of this work has focused on patients with apraxia, a disorder defined by impairments in the production of gestures that cannot be attributed to problems in the actual control of the effectors. With regard to their motor output, many of these patients appear to have lost their knowledge of action semantics; for example, they are unable to pantomime familiar gestures or use tools. Ideomotor

apraxia is characterized as a deficit in the sequencing of skilled actions, resulting in temporal and spatial errors. The individual elements of the action, by themselves, might be correctly performed, but the overriding idea of the action as a whole appears to be lost (Heilman, Maher, Greenwald, & Rothi, 1997). There is some evidence however, that this is not simply a disorder of temporal sequencing but might have to do with loss of action semantics. In a seminal paper, Ochipa, Rothi, and Heilman (1989) reported a case study of a left-handed man who exhibited ideational apraxia following damage to the right hemisphere. Paralleling his difficulty in using tools, the patient showed a severe impairment in action knowledge. For example, he was unable to select the object that was best matched with a particular tool (e.g., a nail with a hammer). Thus his problem was not limited to production. Similarly, his deficit could not be attributed to a language comprehension deficit because he could name the tools and readily pointed to the correct object upon command. Instead, it appeared that the patient had an impairment in action semantics (Ochipa et al., 1989), suggesting that, in part, action semantics and motor representations share forms of representation.

Similarly, conceptual apraxia is defined by a loss of conceptual knowledge about tools and objects associated with gestures. However, unlike ideomotor apraxia, the temporal sequencing of an action is preserved. Conceptual apraxia is characterized by two types of problems: impairments of associative knowledge (tool-action associations such as hammer-pound and tool-object associations such as hammer-nail) and impairments of mechanical knowledge such as the properties and advantages of a given tool. Patients with conceptual apraxia exhibit errors when asked to demonstrate the semantic content of an action (e.g., using a scissor-cutting motion for sawing) even though they are not impaired in object recognition (e.g., they can name the saw). Furthermore, the ability to name the tool associated with a given object is often compromised. These patients tend to have difficulty in associating a tool with the linguistic description of its function (e.g., they are unable to pick out the hammer after hearing the word *pounding*) (Goldenberg & Hagmann, 1998; Heilman et al., 1997). Heilman and colleagues (1997) have argued that this form of apraxia is typically limited to patients with left hemisphere damage, consistent with the hypothesis that there is a linguistic component to the deficit. Moreover, these results further support the hypothesis that there is some degree of overlap between action semantics and nonlanguage motor representations.

Imaging studies provide converging evidence that action semantics is tightly coupled to motor regions. Reading words associated with foot, hand, or mouth actions (e.g., *kick*, *pick*, *lick*) activates premotor areas adjacent to or overlapping with areas that are activated when producing actions with the same effectors (Hauk, Johnsrude, & Pulvermuller, 2004). Tettamanti and his colleagues (2005) found a similar somatotopy in premotor cortex in a sentence comprehension task. Aziz-Zadeh, Wilson, Rizzolatti, and Iacoboni (2006) used an action observation task to localize regions of interest in foot, hand, and mouth regions within premotor cortex. When these participants listened to sentences describing similar actions, these areas were activated in an effector-specific manner. Interestingly, these effects were restricted to the left hemisphere (figure 44.6). This lateralization argues against an account in which the premotor regions are indirectly recruited via imagery given that the video activation patterns had been bilateral. Rather, the left lateralization suggests that premotor cortex is engaged by more abstract linguistic representations.

These data indicate that action semantics and visuomotor representations share common neural substrates. This could explain why, at the behavioral level, interactions between semantics and motor planning are observed. However, we cannot infer causality from patterns of activation in imaging studies; it is unclear whether these premotor activations are essential for action comprehension or reflect some indirect, noncausal recruitment, perhaps via priming from other areas. In the preceding discussion of neurological patients, the locus of the lesions was only crudely specified, generally described in terms of whether the damage was to the left or right hemisphere.

A more direct assessment of the role of motor areas in action comprehension comes from studies in which a more precise specification of the pathology has been described. Action verb comprehension and naming appear to be significantly compromised in patients with motor neuron disease (MND), a neurodegenerative disease with prominent pathology of corticospinal neurons (Bak & Hodges, 1999, 2004; Bak, O'Donovan, Xuereb, Boniface, & Hodges, 2001). Interestingly, these patients do not appear to be impaired in their comprehension and naming of nouns. Postmortem studies indicate that the degeneration in MND patients extends beyond motor and premotor cortex into inferior frontal gyrus (Brodmann areas 44 and 45). Thus, the linguistic problems here might not be directly related to damage to primary and secondary motor regions. Nonetheless, they again demonstrate the neural overlap of regions involved in motor control and action semantics, even when the semantics are accessed linguistically (Bak et al., 2001). Patients with progressive supranuclear palsy (PSP) (Daniele, Giustolisi, Silveri, Colosimo, & Gainotti, 1994) and frontotemporal dementia (Cappa et al., 1998) also show deficits in verb processing, indicating that frontal and frontostriatal brain areas have a role in action semantics.

Perhaps the most comprehensive patient study of action semantics comes from the work of Tranel, Kemmerer, Damasio, Adolphs, and Damasio (2003). Ninety patients with lesions to various brain regions were assessed in their

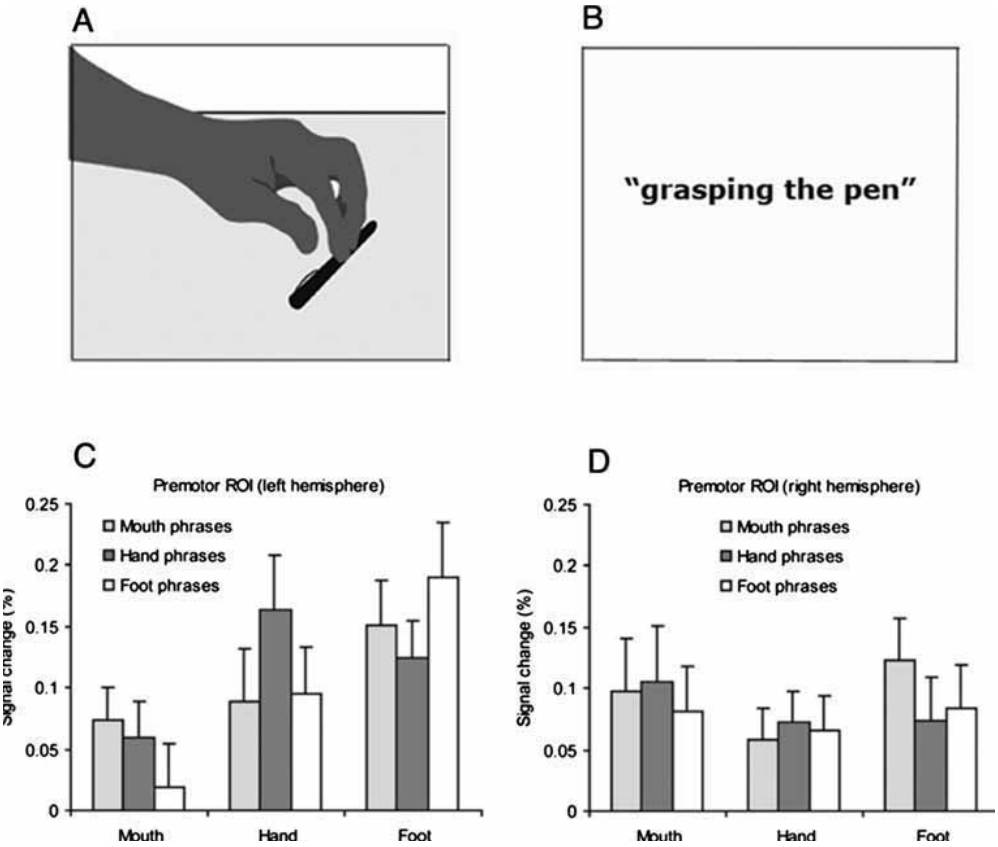


FIGURE 44.6 (*A*) Observation of movements performed by the hand, mouth, or foot was used to localize regions of interests (ROIs) in the premotor cortex. (*B*) The same participants read phrases related to hand, mouth, or foot actions. (*C*) The left hemisphere ROI associated with hand action observation was most active when the participants read phrases described as hand-based actions. The ROI associated with mouth action observation was most active for mouth-related phrases; similarly, the region defined by foot action observation was most active in reading of foot-based actions. (*D*) No significant effects were observed in the right hemisphere ROIs. (Adapted from Aziz-Zadeh et al., 2006.)

ability to retrieve action knowledge, using a task in which the participants matched pictures that depicted related actions. Based on a lesion overlap approach, the highest incidence of impairment was associated with damage to the left premotor/prefrontal cortex, the left parietal region, and the white matter underneath the left posterior middle temporal region. A similar dual pattern of deficit was reported in a study in which aphasic patients were tested for their comprehension of visually or verbally presented actions. Patients with lesions of premotor or parietal area were impaired on these tasks, although lesions in premotor areas were more predictive of the observed deficits (Saygin, Wilson, Dronkers, & Bates, 2004).

Taken together, these studies indicate that action comprehension deficits can be observed in patients who have damage to areas associated with planning actions, in particular premotor and parietal regions. Consistent with the arguments raised in our earlier discussion of hierarchies, these results further challenge a traditional view in which motor control and language are segregated into separate modules, with the latter occupying a supraordinate position that can access the former. Rather, the evidence is more consistent with an embodied cognition framework, one in which our conceptual knowledge of actions is dependent on the systems that are required to produce actions. In the more extreme form, this embodiment would extend to our linguistic knowledge of actions (Feldman, 2006).

Summary

Hierarchy as a word was first used around 1380 to describe the strict relationship between the three layers of angels (seraphim, cherubim, and thrones) ascending toward heaven. Each was subordinate yet dependent on the lower level. In this chapter, we have argued for the existence of a hierarchy in the human brain for organizing complex motor behavior that, like the angels, carries with it distinct functional dependencies. However, unlike the angels, the anatomy of the motor system and the multitude of solutions for achieving complex behaviors suggest that the supraordinate or subordinate roles played by different layers of functional hierarchy can be readily interchanged.

REFERENCES

Albert, N., & Ivry, R. B. (2009). The persistence of spatial interference after extended training in a bimanual drawing task. *Cortex*, *45*, 377–385.

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.*, *16*, 1818–1823.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J. Cogn. Neurosci.*, *19*, 2082–2099.

Bak, T. H., & Hodges, J. R. (1999). Cognition, language and behaviour in motor neurone disease: Evidence of frontotemporal dysfunction. *Dement. Geriatr. Cogn. Disord.*, *10*(Suppl. 1), 29–32.

Bak, T. H., & Hodges, J. R. (2004). The effects of motor neurone disease on language: Further evidence. *Brain Lang.*, *89*, 354–361.

Bak, T. H., O'Donovan, D. G., Xuereb, J. H., Boniface, S., & Hodges, J. R. (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease-dementia-aphasia syndrome. *Brain*, *124*, 103–120.

Bernstein, N. A. (1996). On dexterity and its development. In M. L. Latash & M. T. Turvey (Eds.), *Dexterity and its development*. Mahwah, NJ: Lawrence Erlbaum.

Botvinick, M. M. (2007). Multilevel structure in behaviour and in the brain: A model of Fuster's hierarchy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *362*, 1615–1626.

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.*, *12*, 201–208.

Brinkman, C. (1984). Supplementary motor area of the monkey's cerebral cortex: Short- and long-term deficits after unilateral ablation and the effects of subsequent callosal section. *J. Neurosci.*, *4*, 918–929.

Cappa, S. F., Binetti, G., Pezzini, A., Padovani, A., Rozzini, L., & Trabucchi, M. (1998). Object and action naming in Alzheimer's disease and frontotemporal dementia [see comment]. *Neurology*, *50*, 351–355.

Cooper, R., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychol. Rev.*, *113*, 887–916; discussion, 917–831.

Culham, J. C., Danckert, S. L., DeSouza, J. F., Gati, J. S., Menon, R. S., & Goodale, M. A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Exp. Brain Res.*, *153*, 180–189.

d'Avella, A., & Bizzi, E. (2005). Shared and specific muscle synergies in natural motor behaviors. *Proc. Natl. Acad. Sci. USA*, *102*, 3076–3081.

Daniele, A., Giustolisi, L., Silveri, M. C., Colosimo, C., & Gainotti, G. (1994). Evidence for a possible neuroanatomical basis for lexical processing of nouns and verbs. *Neuropsychologia*, *32*, 1325–1341.

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends Cogn. Sci.*, *4*, 423–431.

Diedrichsen, J., Grafton, S., Albert, N., Hazeltine, E., & Ivry, R. B. (2006). Goal-selection and movement-related conflict during bimanual reaching movements. *Cereb. Cortex*, *16*, 1729–1738.

Diedrichsen, J., Hazeltine, E., Kennerley, S., & Ivry, R. B. (2001). Moving to directly cued locations abolishes spatial interference during bimanual actions. *Psychol. Sci.*, *12*, 493–498.

Dum, R. P., & Strick, P. L. (1991). The origin of corticospinal projections from the premotor areas in the frontal lobe. *J. Neurosci.*, *11*, 667–689.

Dum, R. P., & Strick, P. L. (1996). Spinal cord terminations of the medial wall motor areas in macaque monkeys. *J. Neurosci.*, *16*, 6513–6525.

Dum, R. P., & Strick, P. L. (2005). Frontal lobe inputs to the digit representations of the motor areas on the lateral surface of the hemisphere. *J. Neurosci.*, *25*, 1375–1386.

Feldman, J. A. (2006). *From molecule to metaphor: A neural theory of language*. Cambridge, MA: MIT Press.

Franz, E. A., Eliasson, A. C., Ivry, R. B., & Gazzaniga, M. S. (1996). Dissociation of spatial and temporal coupling in the bimanual movements of callosotomy patients. *Psychol. Sci.*, *7*, 306–310.

Franz, E. A., Zelaznik, H. N., Swinnen, S. S., & Walter, C. (2001). Spatial conceptual influences on the coordination of bimanual actions: When a dual task becomes a single task. *J. Motor Behav.*, *33*, 103–112.

Frey, S. H. (2008). Tool use, communicative gesture and cerebral asymmetries in the modern human brain. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *363*, 1951–1957.

Fuster, J. M. (1995). *Memory in the cerebral cortex*. Cambridge, MA: MIT Press.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in reason and language. *Cogn. Neuropsychol.*, *22*, 455–479.

Gentilucci, M., Benuzzi, F., Bertolani, L., Daprati, E., & Gangitano, M. (2000). Language and motor control. *Exp. Brain Res.*, *133*, 468–490.

Gentilucci, M., & Gangitano, M. (1998). Influence of automatic word reading on motor control. *Eur. J. Neurosci.*, *10*, 752–756.

Giszter, S. F., Mussa-Ivaldi, F. A., & Bizzi, E. (1993). Convergent force fields organized in the frog's spinal cord. *J. Neurosci.*, *13*, 467–491.

Glover, S., & Dixon, P. (2002). Semantics affect the planning but not control of grasping. *Exp. Brain Res.*, *146*, 383–387.

Goldenberg, G., & Hagmann, S. (1998). Tool use and mechanical problem solving in apraxia. *Neuropsychologia*, *36*, 581–589.

Goodale, M. A., Meenan, J. P., Bulthoff, H. H., Nicolle, D. A., Murphy, K. J., & Racicot, C. I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. *Curr. Biol.*, *4*, 604–610.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, *349*, 154–156.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1998). Abstract and effector-specific representations of motor sequences identified with PET. *J. Neurosci.*, *18*, 9420–9428.

Grafton, S. T., Salidis, J., & Willingham, D. B. (2001). Motor learning of compatible and incompatible visuomotor maps. *J. Cogn. Neurosci.*, *13*, 217–231.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychol. (Amst.)*, *107*, 293–321.

Haggard, P., & Wing, A. (1997). On the hand transport component of prehensile movements. *J. Motor Behav.*, *29*, 282–287.

Hamilton, A. F., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *J. Neurosci.*, *26*, 1133–1137.

Hamilton, A. F., de C., & Grafton, S. T. (2007). The motor hierarchy: From kinematics to goals and intentions. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition. Attention and performance: Vol. 22*. (pp. 381–408). Oxford, UK: Oxford University Press.

Hamilton, A. F., & Grafton, S. T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cereb. Cortex*, *18*, 1160–1168.

Hartmann, K., Goldenberg, G., Daumüller, M., & Hermsdörfer, J. (2005). It takes the whole brain to make a cup of coffee: The neuropsychology of naturalistic actions involving technical devices. *Neuropsychologia*, *43*, 625–637.

Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.

Heilman, K. M., Maher, L. M., Greenwald, M. L., & Rothi, L. J. (1997). Conceptual apraxia from lateralized lesions. *Neurology*, *49*, 457–464.

Heuer, H., Kleinsorge, T., Spijkers, W., & Steglich, W. (2001). Static and phasic cross-talk effects in discrete bimanual reversal movements. *J. Motor Behav.*, *33*, 67–85.

Ivry, R. B., Diedrichsen, J., Spencer, R. C. M., Hazeltine, E., & Semjen, A. (2004). A cognitive neuroscience perspective on bimanual coordination. In S. Swinnen & J. Duysens (Eds.), *Neurobehavioral determinants of interlimb coordination* (pp. 259–295). Boston: Kluwer.

Ivry, R. B., & Robertson, L. C. (1998). *The two sides of perception*. Cambridge, MA: MIT Press.

Jeannerod, M. (1984). The timing of natural prehension movements. *J. Motor Behav.*, *16*, 235–254.

Jeannerod, M. (1986). The formation of finger grip during prehension: A cortically mediated visuomotor pattern. *Behav. Brain Res.*, *19*, 99–116.

Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford, UK: Blackwell.

Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. *Trends Neurosci.*, *18*, 314–320.

Johnson, S. H., & Grafton, S. T. (2003). From "acting on" to "acting with": The functional anatomy of object-oriented action schemata. *Prog. Brain Res.*, *142*, 127–139.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., et al. (1998). The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proc. Natl. Acad. Sci. USA*, *95*, 861–868.

Keele, S. W., Cohen, A., & Ivry, R. (1990). Motor programs: Concepts and issues. In M. Jeannerod (Ed.), *Attention and performance: Vol. 13. Motor representation and control* (pp. 77–110). Hillsdale, NJ: Lawrence Erlbaum.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *J. Neurosci.*, *20*, 3310–3318.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

Lu, X., & Ashe, J. (2005). Anticipatory activity in primary motor cortex codes memorized movement sequences. *Neuron*, *45*, 967–973.

Matsuzaka, Y., Picard, N., & Strick, P. L. (2007). Skill representation in the primary motor cortex after long-term practice. *J. Neurophysiol.*, *97*, 1819–1832.

Mechsner, F., Kerzel, D., Knoblich, G., & Prinz, W. (2001). Perceptual basis of bimanual coordination. *Nature*, *414*, 69–73.

Ochipa, C., Rothi, L. J., & Heilman, K. M. (1989). Ideational apraxia: A deficit in tool selection and use. *Ann. Neurol.*, *25*, 190–193.

Rice, N. J., Tunik, E., & Grafton, S. T. (2006). The anterior intraparietal sulcus mediates grasp execution, independent of requirement to update: New insights from transcranial magnetic stimulation. *J. Neurosci.*, *26*, 8176–8182.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron*, *31*, 889–901.

Rizzolatti, G., & Matelli, M. (2003). Two different streams form the dorsal visual system: Anatomy and functions. *Exp. Brain Res.*, *153*, 146–157.

Roland, P. E., Larsen, B., Lassen, N. A., & Skinhoj, E. (1980). Supplementary motor area and other cortical areas in organization of voluntary movements in man. *J. Neurophysiol.*, *43*, 118–136.

Roland, P. E., Skinhøj, E., Lassen, N. A., & Larsen, B. (1980). Different cortical areas in man in organization of voluntary movements in extrapersonal space. *J. Neurophysiol.*, *43*, 137–150.

Rosenbaum, D. A., Meulenbroek, R. G., & Vaughan, J. (2001). Planning reaching and grasping movements: Theoretical premises and practical implications. *Motor Control*, *5*, 99–115.

Rosenbaum, D. A., Vaughan, J., Barnes, H. J., & Jorgensen, M. J. (1992). Time course of movement planning: Selection of handgrips for object manipulation. *J. Exp. Psychol. Learn. Mem. Cogn.*, *18*, 1058–1073.

Rumiati, R. I., & Tessari, A. (2002). Imitation of novel and well-known actions: The role of short-term memory. *Exp. Brain Res.*, *142*, 425–433.

Rumiati, R. I., Weiss, P. H., Tessari, A., Assmus, A., Zilles, K., Herzog, H., et al. (2005). Common and differential neural mechanisms supporting imitation of meaningful and meaningless actions. *J. Cogn. Neurosci.*, *17*, 1420–1431.

Sadato, N., Yonekura, Y., Waki, A., Yamada, H., & Ishii, Y. (1997). Role of the supplementary motor area and the right premotor cortex in the coordination of bimanual finger movements. *J. Neurosci.*, *17*, 9667–9674.

Saygin, A. P., Wilson, S. M., Dronkers, N. F., & Bates, E. (2004). Action comprehension in aphasia: Linguistic and non-linguistic deficits and their lesion correlates. *Neuropsychologia*, *42*, 1788–1804.

Schöner, G., & Kelso, J. A. (1988). Dynamic pattern generation in behavioral and neural systems. *Science*, *239*, 1513–1520.

Semjen, A., & Ivry, R. B. (2001). The coupled oscillator model of between-hand coordination in alternate-hand tapping: A reappraisal. *J. Exp. Psychol. Hum. Percept. Perform.*, *27*, 251–265.

Spencer, R. C. M., Semjen, A., Yang, S., & Ivry, R. B. (2007). An event-based account of coordination stability. *Psychon. Bull. & Rev.*, *13*, 702–710.

Tessari, A., & Rumiati, R. I. (2004). The strategic control of multiple routes in imitation of actions. *J. Exp. Psychol. Hum. Percept. Perform.*, *30*, 1107–1116.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.*, *17*, 273–281.

Tranel, D., Kemmerer, D., Damasio, H., Adolphs, R., & Damasio, A. R. (2003). Neural correlates of conceptual knowledge for actions. *Cogn. Neuropsychol.*, *20*, 409–432.

Tunik, E., Frey, S. H., & Grafton, S. T. (2005). Virtual lesions of the anterior intraparietal area disrupt goal-dependent on-line adjustments of grasp. *Nat. Neurosci.*, *8*, 505–511.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1882.

Wolpert, D. M., Goodbody, S. J., & Husain, M. (1998). Maintaining internal representations: The role of the human superior parietal lobe. *Nat. Neurosci.*, *1*, 529–533.

VI

MEMORY

Chapter 45 SUZUKI 659

46 SHRAGER AND SQUIRE 675

47 NADER 691

48 RACE, KUHL, BADRE, AND WAGNER 705

49 KENSINGER 725

50 MILLER 739

51 SCHACTER, ADDIS, AND BUCKNER 751

Introduction

DANIEL L. SCHACTER

MEMORY IS so fundamental to the operation of the brain and mind that students of the topic could be forgiven for feeling that their object of study is perhaps the most central in all of cognitive neuroscience. But what is memory? As I wrote in the introductory section to the memory chapters in the 2005 edition of this book, we cannot yet provide a satisfactory answer to this question, but we do know something worth knowing about what memory is *not*: it is not a single entity or concept. Indeed, as Endel Tulving stated in his introduction to the memory chapters in this volume's first edition, "Memory is many things, even if not everything that has been labeled memory corresponds to what cognitive neuroscientists think of as memory. Memory is a gift of nature, the ability of living organisms to retain and to utilize acquired information. The term is closely related to *learning*, in that memory in biological systems always entails learning (the acquisition of information) and in that learning implies retention (memory) of such information." Tulving also characterized memory as a trick of evolution, a biological abstraction, and a convenient chapter heading for certain kinds of problems that scientists study.

Tulving's observations still make a great deal of sense. For example, the starting point for virtually any scientific analysis of memory involves a decomposition into processes of encoding, storage and consolidation, and retrieval. Furthermore, a prominent theme in cognitive neuroscience for the past two decades has been that memory can be divided into multiple forms or systems—collections of processes that operate on different kinds of information and according to different rules. Forms of memory such as working, episodic, semantic, priming, and procedural memory are all familiar to contemporary researchers.

The idea that memory is not a single thing also extends to memory's imperfections. Memory is not a simple matter

of success or failure. Memory can go awry because of either forgetting or distortion, and each of these types of error can be subdivided into several distinguishable forms. Furthermore, memory interacts in important ways with a number of related processes, including emotion, cognitive control, and planning. A full understanding of memory requires us to address its varied manifestations and complexities. Happily, memory researchers have risen to the challenge by exploring and perhaps beginning to unravel the many complexities of memory.

The seven chapters in this section highlight many facets of memory at different levels of analysis. The starting point for cognitive neuroscience analyses of memory typically begins in the medial temporal lobe. Ever since the groundbreaking observations of Scoville and Milner during the 1950s concerning the severe amnesia observed in patient HM after bilateral resection of the medial temporal lobe for relief of intractable epilepsy, attempting to understand this region's role in memory and learning has constituted a kind of holy grail for memory researchers. It is therefore appropriate that the section begins with Suzuki's chapter concerning the neuroanatomy of the medial temporal lobe (chapter 45). She provides a detailed analysis that focuses in particular on similarities and differences across species in a number of key medial temporal lobe structures, including perirhinal, parahippocampal, and entorhinal cortices. Among the interesting differences observed, Suzuki notes that in monkey perirhinal cortex, unimodal input is dominated by the visual modality, whereas the rat perirhinal cortex receives a wider range of inputs from all sensory modalities. Suzuki delineates possibly important functional consequences of such differences.

In chapter 46, Shrager and Squire focus on work with human amnesic patients to examine spared and impaired functions after medial temporal lobe damage. They report a series of refined experimental studies concerning amnesic patients who have well-characterized lesions in order to address a number of topics that have been central to recent discussions of medial temporal lobe function, including working memory, habit learning, recollection versus familiarity, path integration, remote memory, and conscious awareness. Their observations help to delineate the role of the medial temporal lobe in each of the foregoing aspects of memory and learning.

Nader in chapter 47 focuses on the recently rediscovered phenomenon of reconsolidation, one of the most intensively investigated and hotly debated topics in neuroscience-based memory research during the past decade. It has long been accepted that memory is a time-dependent process, involving a consolidation phase where new memories are initially unstable or labile, and then over time become more stable and resistant to disruption. Reconsolidation refers to the observation that reactivating a seemingly consolidated memory can, under a number of conditions, return it transiently to a labile state in which it is again subject to disruption. Nader reviews recent experimental evidence, considers various alternative interpretations of the phenomenon, and attempts to link reconsolidation with approaches to memory that emphasize its constructive nature.

In chapter 48, Race, Kuhl, Badre, and Wagner examine the interface between memory and cognitive control processes, which guide thought and action in accordance with current goals. They review fMRI studies concerned with the contributions of specific regions within the frontal lobe to cognitive control over memory, focusing especially on situations in which competition between memories creates interference, and where ineffective retrieval cues yield uncertainty. Their discussion of the theoretical implications of dissociations between specific frontal subregions in several task domains illustrates the impressive specificity of neuroanatomical and functional conclusions that can be drawn on the basis of imaging studies.

Just as the interface between memory and cognitive control has brought a major topic of experimental and theoretical concerns to the surface, so has the interface between memory and emotion. Kensinger in chapter 49 reviews this increasingly impressive body of research, which shows that interactions with emotion can arise at every phase of the memory process, including encoding, consolidation, and retrieval. She considers neuroimaging data that clarify the role of the amygdala in emotional memory and that reveal conditions under which amygdala activity corresponds with increased accuracy of emotional memories. Kensinger also considers how individual differences can influence emotion-memory interactions.

The latter topic constitutes the central focus of Miller's chapter 50, especially in relation to neuroimaging studies of memory functions. Though individual differences in memory have often been overlooked in memory research, Miller argues that for neuroimaging studies especially, group analyses can be incomplete and even misleading. Miller illustrates this point with recent work from his own laboratory, and he also integrates observations concerning individual differences with some key theoretical issues in memory research.

Though the preceding six chapters cover many different facts of memory, they all approach memory as a process that is concerned with recovering information from the past. In the concluding chapter of this part (chapter 51), Schacter, Addis, and Buckner consider a recent and rapidly evolving literature that implicates memory as a key player in allowing individuals to think about and simulate possible happenings in the future. They discuss striking observations from neuroimaging and neuropsychological studies showing cognitive and neural overlap between the processes involved in remembering the past and imagining the future, and relate

these observations to the idea that memory is a fundamentally constructive process, sometimes prone to errors and illusions. They consider the possibility that the flexible use of information from memory to simulate alternative future scenarios constitutes a key function of a constructive memory system.

The chapters in this section reveal expansions in both the depth and breadth of memory research, which bodes well for the future of the enterprise. We cannot know with any certainty what path memory research will follow in the upcoming years, but we can be confident that it will be exciting to find out.

45 Comparative Analysis of the Cortical Afferents, Intrinsic Projections, and Interconnections of the Parahippocampal Region in Monkeys and Rats

WENDY A. SUZUKI

ABSTRACT Detailed neuroanatomical studies that focused on the connections of the medial temporal lobe in monkeys provided critical clues toward identifying the structures important for normal declarative memory. These structures include the hippocampus together with the surrounding and strongly interconnected entorhinal, perirhinal, and parahippocampal cortices. Detailed anatomical descriptions of the connections of the analogous cortical regions in rats suggest both similarities and differences in the connections of these cortical medial temporal lobe areas across species. In this chapter we will review the quantitative anatomical studies describing the cortical inputs, intrinsic projections, and interconnections of the entorhinal, perirhinal, and parahippocampal cortices in monkeys and rats. A detailed understanding of the cross-species similarities and differences in the anatomical organization of these regions can provide valuable insight into understanding the core mnemonic functions of these areas.

The landmark description by Scoville and Milner (1957) of a group of brain-damaged patients including the well-known amnesic patient HM demonstrated for the first time that bilateral damage limited to the region of the medial temporal lobe in humans resulted in a permanent and devastating memory impairment. Later studies showed that patients with medial temporal lobe damage exhibited a memory loss that was selective for fact and event memory (i.e., declarative memory; Eichenbaum & Cohen, 2001; Gabrieli, 1998; Squire, Knowlton, & Musen, 1993; Squire, 1992). While the original report of Scoville and Milner (1957) first identified the region of the medial temporal lobe as key for memory function, a convergence of systematic anatomical and neurobehavioral studies in animal model systems was critical for identifying which of the structures in the medial temporal lobe, when damaged, were responsible for the severe declarative memory impairment seen in patient HM.

Whereas early experimental lesion (Mishkin, 1978) and neurophysiology studies (O'Keefe & Nadel, 1978) tended to focus on the role of the hippocampus in declarative-like memory, because later studies showed that selective hippocampal lesions in humans (Zola-Morgan, Squire, & Amaral, 1986) resulted in mild memory impairment relative to the impairment seen in patient HM, this finding suggested that brain areas beyond the hippocampus may also be involved. The anatomical studies of Amaral and colleagues (Amaral, Insausti, & Cowan, 1987; Insausti, Amaral, & Cowan, 1987a, 1987b) provided critical insight into which other medial temporal lobe areas might be participating in declarative memory. Specifically, their quantitative neuroanatomical studies showed that the monkey entorhinal cortex, the major source of cortical inputs to the hippocampus, received the vast majority of its cortical projections from the surrounding perirhinal and parahippocampal cortices. Further anatomical studies revealed that the perirhinal and parahippocampal cortices (Suzuki & Amaral, 1994a, 1994b) received a powerful convergence of unimodal and polymodal cortical inputs and in this way served as a critical relay for multimodal information into the hippocampal formation (i.e., hippocampus and entorhinal cortex). Taken together, these anatomical insights were critical in focusing attention on the possible mnemonic role of the entorhinal, perirhinal, and parahippocampal cortices. A convergence of subsequent lesion studies (Leonard, Amaral, Squire, & Zola-Morgan, 1995; Suzuki, Zola-Morgan, Squire, & Amaral, 1993; Meunier, Bachevalier, Mishkin, & Murray, 1993; Zola-Morgan, Squire, Amaral, & Suzuki, 1989; Murray &

WENDY A. SUZUKI Center for Neural Science, New York University, New York, New York

Mishkin, 1986) and neurophysiological studies (Miller, Li, & Desimone, 1993, 1991; Riches, Wilson, & Brown, 1991; Baylis & Rolls, 1987; Brown, Wilson, & Riches, 1987) in monkeys together with findings from brain-damaged patients with detailed MRI localization (Stefanacci, Buffalo, Schmolck, & Squire, 2000; Corkin, Amaral, Gilberto Gonzalez, Johnson, & Hyman, 1997) confirmed the critical contribution of these cortical medial temporal lobe areas to declarative memory.

Parallel neuroanatomical, lesion, and neurophysiological studies in rodents have also made important progress in defining the functional organization of these cortical medial temporal lobe regions. Anatomically, the perirhinal and parahippocampal cortices in monkeys are thought to be homologous to the perirhinal and postrhinal cortices in rats, respectively (Burwell & Witter, 2002; Burwell & Amaral, 1998a; Burwell, Witter, & Amaral, 1995). Neuroanatomical studies revealed that similar to monkeys, these cortical areas in rats also receive a strong convergence of multimodal sensory information (Burwell & Amaral, 1998a; Deacon, Eichenbaum, Rosenberg, & Eckmann, 1983). Lesion and neurophysiology studies, however, have revealed both similarities and differences in the functional organization of these cortical areas across species. Behavioral neurophysiology studies support the idea that like the monkey perirhinal cortex (Miller et al., 1993, 1991; Riches et al., 1991; Baylis & Rolls, 1987; Brown et al., 1987), the rat perirhinal cortex signals visual recognition by responding significantly less upon stimulus repetition (Xiang & Brown, 1998; Brown & Xiang, 1998). However, recent lesion studies focused on the rat perirhinal cortex implicate this region in context memory (Bucci, Saddoris, & Burwell, 2002; Bucci, Phillips, & Burwell, 2000), a function that has not been examined in the monkey perirhinal cortex. Lesion (Malkova & Mishkin, 2003) and physiology studies (Rolls & Xiang, 2005; Cahusac, Miyashita, & Rolls, 1989) focused on the monkey parahippocampal cortex have implicated this region in spatial memory. Lesion studies in rats have implicated the postrhinal cortex in context memory (Bucci et al., 2002, 2000), but findings on tasks of spatial memory have been mixed (Burwell, Bucci, Sanborn, & Jutras, 2004; Liu & Bilkey, 2002). Physiology studies reported positional firing characteristics in rat postrhinal cells that differed from those previously seen in the rodent hippocampus or medial entorhinal cortex (Fyhn, Molden, Witter, Moser, & Moser, 2004; Burwell & Hafeman, 2003; Shapiro, Tanila, & Eichenbaum, 1997).

It is clear from this brief review that while there are similarities in the functional organization of these cortical areas across monkeys and rats, differences are emerging as well. Given the reliance on these animal model systems for understanding human memory function, it is important to accurately identify both the core mnemonic and processing functions of these areas that are similar across species and the species-specific differences. In particular, the recent discovery of grid cells in the rat medial entorhinal cortex (Hafting, Fyhn, Bonnevie, Moser, & Moser, 2008; Fyhn, Hafting, Treves, Moser, & Moser, 2007; Hafting, Fyhn, Molden, Moser, & Moser, 2005; Fyhn et al., 2004) raises fascinating new questions about the homology of physiological findings across these cortical medial temporal lobe areas in monkeys and rats. Grid cells in the rodent medial entorhinal cortex exhibit a striking pattern of spatially selective firing in that they represent large extents of the physical environment in a gridlike pattern of tessellating equilateral triangles. Could there be gridlike cells in the monkey entorhinal cortex as well? One way to address the general question of the homology of the cortical medial temporal lobe areas across monkeys and rats is to take advantage of the growth over the last 10 years of detailed and quantitative neuroanatomical studies of these cortical areas in both species. There is now a critical mass of such anatomical data that allows more detailed cross-species comparisons of the connections of the cortical areas surrounding the hippocampus. In this chapter we will characterize the flow of cortical information into the perirhinal, parahippocampal, postrhinal, and entorhinal cortices together with the prominent intrinsic projections and interconnections of these areas in both monkeys and rats.

Boundaries and nomenclature of the perirhinal, parahippocampal, postrhinal, and entorhinal cortices: Many controversies and some agreement

The Parahippocampal Region in Monkeys

The perirhinal, parahippocampal, and entorhinal cortices are often referred to as the *parahippocampal region* (Burwell & Witter, 2002). In monkeys, these areas are situated on the anterior-ventral portion of the medial temporal lobe (figure 45.1*A*) and surround the amygdala anteriorly and the hippocampus posteriorly. While the boundaries and nomenclature of the monkey entorhinal cortex are generally accepted, substantial controversy currently exists over the nomenclature and precise boundaries of both the perirhinal and parahippocampal cortices. Although a detailed discussion of the source of these controversies is beyond the scope of this chapter, we briefly mention the boundaries that are most disputed in both the monkey and the rat literature.

The monkey perirhinal cortex is situated lateral to the fundus of the rhinal sulcus and extends from the posterior border of the rhinal sulcus on the ventral surface of the temporal lobe to the anterior and dorsal portion of the rhinal sulcus on the temporal pole (figure 45.1*A*). It is bordered medially by the entorhinal cortex, laterally by visual area TE, and posteriorly by the parahippocampal cortex. The perirhi-

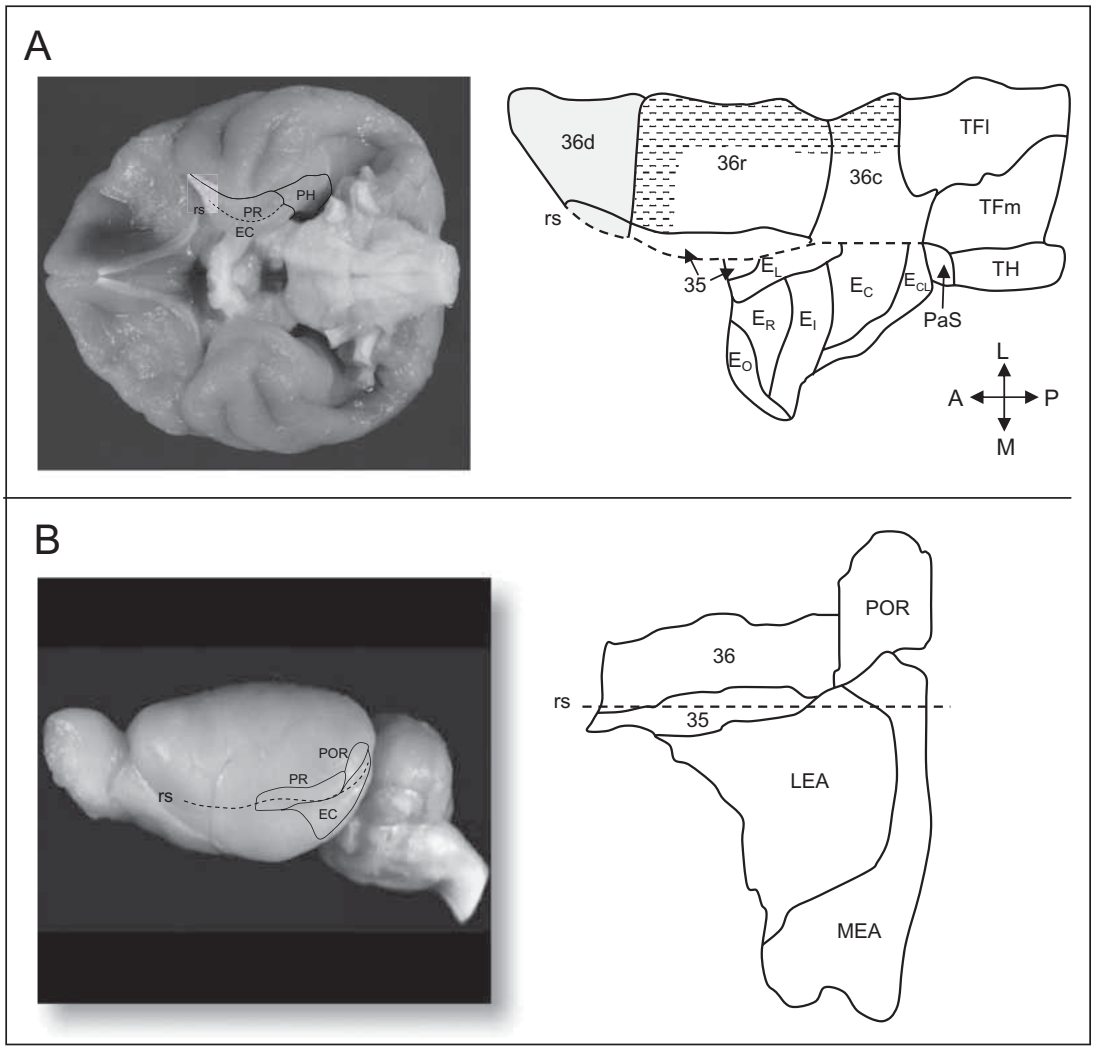


FIGURE 45.1 (*A*) *Left:* Photograph illustrating the ventral view of the macaque monkey brain showing the locations of the entorhinal (EC), perirhinal (PR), and parahippocampal (PH) cortices surrounding the rhinal sulcus (rs). The shaded region at the level of the dorsal temporal pole corresponds to area 36d. *Right:* An unfolded representation of the same cortical medial temporal lobe areas shown on the left is illustrated along with major subdivisions. The perirhinal cortex is subdivided into areas 35, 36r, and 36c. The parahippocampal cortex includes areas TFl, TFm, and TH. The monkey entorhinal cortex is subdivided into the olfactory (E_O), rostral (E_R), lateral (E_L), intermediate (E_I), caudal (E_C), and caudal limited (E_{CL}) subdivisions. The location of area 36d is indicated on the unfolded map but not included within the boundaries of the perirhinal cortex for this chapter (see text for explanation). The stippled region within areas 36r and 36c shows the approximate extent and location of the disputed anterior and lateral borders of the perirhinal cortex. We will use the more anterior and lateral boundaries of the perirhinal cortex as described by Suzuki and Amaral (1994b, 2003a). (*B*) *Left:* Photograph illustrating the lateral view of a rat brain showing the locations of the entorhinal (EC), perirhinal (PR), and postrhinal (POR) cortices. *Right:* An illustration of an unfolded representation of the same cortical areas along with all major subdivisions. As in the monkey, the rat perirhinal cortex is subdivided in areas 35 and 36 while the entorhinal cortex is subdivided into the lateral entorhinal area (LEA) and the medial entorhinal area (MEA). The rat postrhinal cortex has not been subdivided further. Additional abbreviations: PaS, parasubiculum; A, anterior; P, posterior; M, medial; L, lateral.

nal cortex has been further subdivided into area 35, which forms a long and narrow strip of cortex situated in the fundus and lateral bank of the rhinal sulcus, and a larger, more laterally situated area 36. Area 36 has further been subdivided into two major subdivisions (areas 36r and 36c) based on cytoarchitectonic criteria. Two main controversies exist over the borders of the monkey perirhinal cortex. One concerns whether the cortex of the temporal pole adjacent to the rhinal sulcus should also be considered part of the perirhinal cortex (Saleem, Price, & Haskikawa, 2007; Kondo, Saleem, & Price, 2005, 2003; Suzuki & Amaral, 1994a; Insausti et al., 1987a). The disputed regions, illustrated in the unfolded map in

figure 45.1*A*, include both area 36d and the anterior stippled region in area 36r that represents the ventral portion of the temporal pole. The second controversy concerns the precise location of the lateral border of area 36 with area TE (Saleem et al., 2007; Kondo et al., 2005, 2003; Suzuki & Amaral, 1994a; Insausti et al., 1987a). This disputed area is shown as the lateral stippled region in areas 36r and 36c in the unfolded map of figure 45.1*A*. In this chapter, we will focus on the connections of the portion of the perirhinal cortex situated on the ventral surface of the brain, which includes part of the ventral temporal pole that has connections and cytoarchitectonic features similar to the ventral perirhinal areas (Kondo et al., 2005; Suzuki & Amaral, 2003a, 2003b, 1994a), as well as the more laterally situated border of area 36 with area TE as described by Suzuki and Amaral (2003a, 1994a). This includes both areas 36r and 36c illustrated in figure 45.1*A*. However, we exclude area 36d of Suzuki and Amaral (1994a), since several studies have shown that this region has different connections from the rest of the ventrally situated perirhinal cortex (Saleem, Kondo, & Price, 2008; Kondo et al., 2003; Suzuki & Amaral, 1994a).

The monkey parahippocampal cortex is situated just caudal to both the perirhinal and entorhinal cortices (figure 45.1*A*). This cortical area is made up of areas TH and TF, and area TF contains two subdivisions defined by cytoarchitectonic features (areas TFm and TFl). Both Blatt, Pandya, & Rosene (2003) and Suzuki and Amaral (2003a, 1994a) place the lateral border of area TF at the level of occipitotemporal sulcus. However, Price and colleagues (Saleem et al., 2007; Kondo et al., 2005) have placed this lateral border approximately 3 mm more medial at approximately the level of Suzuki and Amaral's area TFm. In this review we will use the nomenclature and boundaries for the parahippocampal cortex described by Suzuki and Amaral (2003a, 1994a).

The third component of the parahippocampal region is the entorhinal cortex located in the ventromedial part of the rostral third of the temporal lobe (figure 45.1*A*). The entorhinal cortex has further been subdivided into six subdivisions based mainly on cytoarchitectonic criteria (Amaral et al., 1987). In general, the monkey entorhinal cortex enjoys the least controversy over the location of its borders, compared to the perirhinal and parahippocampal cortices.

The Parahippocampal Region in Rats In rats the parahippocampal region occupies a location relative to the rhinal sulcus that is similar to that of monkeys with some subtle differences. Similar to monkeys, the perirhinal cortex in rats comprises areas 35 and 36. These two subdivisions form two strips of cortex associated with the mid- to posterior portions of the rhinal sulcus (figure 45.1*B*). Area 36 is a wider strip of the cortex that occupies most of the dorsal bank of the rhinal sulcus and part of the laterally adjacent cortex. Area 35 is situated medially and occupies the ventral bank and fundus of the rhinal sulcus. Similar to the monkey, both the anterior border of the rat perirhinal cortex with the insula and the lateral border with area TEv have both varied widely in the literature (Burwell et al., 1995). Here we will use the borders described by Burwell and colleagues (Burwell, 2001).

In contrast to the perirhinal cortex, which shares substantial cytoarchitectonic features between the rat and monkey brain, the postrhinal cortex in rats and the parahippocampal cortex in monkeys do not share clear cytoarchitectonic features. For this reason, Burwell and colleagues (1995) called the region the postrhinal cortex following the earlier rodent literature (Burwell et al., 1995; Deacon et al., 1983) rather than using the nomenclature used in monkeys (Van Hoesen, Pandya, & Butters, 1975; Van Hoesen & Pandya, 1975a, 1975b). While cytoarchitectonic features are not shared across the rat postrhinal and monkey parahippocampal cortex, as will be described in detail later, they exhibit similarities in their connections that have led to the hypothesis that these two areas are homologous (Burwell, 2000; Burwell & Amaral, 1998a).

In rats, the entorhinal cortex is situated on the ventral and caudal surface of the brain medial to the posterior portion of the rhinal sulcus (figure 45.1*B*). At anterior levels its lateral border is situated medial to the rhinal sulcus, and at more caudal levels the entorhinal cortex extends laterally past the fundus of the rhinal sulcus. The rat entorhinal cortex has historically been subdivided into a lateral entorhinal area (LEA) and a medial entorhinal area (MEA) based on morphological criteria (Blackstad, 1956; Krieg, 1946a, 1946b).

Connections of the perirhinal, parahippocampal/postrhinal, and entorhinal cortex: Afferent, intrinsic, and interconnections

The Monkey Perirhinal Cortex

Cortical afferents Early anatomical studies first identified the monkey perirhinal cortex as receiving convergent input from areas involved in processing multiple sensory modalities (Martin-Elkins & Horel, 1992; Seltzer & Pandya, 1976; Van Hoesen et al., 1975; Van Hoesen & Pandya, 1975a, 1975b; Jones & Powell, 1970). More recent retrograde and anterograde tract-tracing studies have provided more detailed and quantitative description of these convergent inputs (Saleem et al., 2008; Kondo et al., 2005; Saleem & Tanaka, 1996; Rockland, Saleem, & Tanaka, 1994; Suzuki & Amaral, 1994a). The single most prominent input to perirhinal area 36 arises from the anterior and medial portions of unimodal visual area TE, a major component of the "ventral visual pathway" important for processing object information (Ungerleider & Mishkin, 1982). The same

regions of medial area TE converge on all parts of the area 36. The second strongest input to area 36 arises from area TF of the adjacent parahippocampal cortex with the strongest inputs originating from the anterior two-thirds of area TF and terminating throughout area 36. Only weak projections are seen from area TH. Moderate projections arise from the visual areas of the ventral bank of the superior temporal sulcus (STSv) that terminate anteriorly in the perirhinal cortex. Weak projections from the polymodal areas of the dorsal bank of the STS (STSd) and area 36d also terminate anteriorly in area 36. Weak projections from orbital frontal areas 11, 12, and 13 (Kondo et al., 2005), insular cortex, and anterior cingulate cortex terminate throughout area 36. In general area 35 receives a pattern of cortical inputs similar to that of area 36, though it receives a relatively stronger input from the dorsal bank of the superior temporal sulcus (see box 45.1).

Intrinsic projections The perirhinal cortex also exhibits prominent intrinsic projections such that each major perirhinal subdivision (36r, 36c, and 35) has prominent interconnections within that subdivision and moderate projections to the other subdivisions (shading in figure 45.2*A*; Lavenex, Suzuki, & Amaral, 2004). Given the strong convergence of projections from areas TE and the parahippocampal cortex to all levels of the perirhinal cortex, this observation suggests that

Box 45.1 Calculations of the Relative Strength of Cortical Inputs

To accurately and systematically quantify the relative strength of the cortical inputs shown in figures 45.2–45.4, we used data derived from retrograde labeling studies in both monkeys and rats. For the cortical inputs to the monkey perirhinal and parahippocampal cortices, we used the quantitative retrograde tracer data from table 1 of Suzuki and Amaral (1994a). For the cortical inputs to the monkey entorhinal cortex, we used the retrograde data from table 1 of Insausti, Amaral, and Cowan (1987a). For the relative proportions of cortical inputs to the rat perirhinal and postrhinal cortex, we used retrograde data from Furtak, Wei, Agster, and Burwell (2007), and for the cortical inputs to the rat entorhinal cortex, we used the retrograde data from Kerr, Agster, Furtak, & Burwell (2007). Based on the number or proportion of retrogradely labeled cells reported in these studies, we next defined the strongest afferent projections (shown with the largest lettering in figures 45.2–45.4) as those projections that represented 20% or more of the total retrogradely labeled cells in cortical areas. Medium-sized afferent projections shown with medium-sized lettering represent between 11% and 19% of total cortical labeling, and weak afferent projections shown with the smallest lettering represent 10% or less of the cortical inputs. To complete these calculations, we also recalibrated the retrograde data presented in both Furtak and colleagues (2007) and Kerr and colleagues (2007) in the following way. First, we recalculated the relative proportion of cortical inputs to the rat perirhinal and postrhinal cortices to include inputs from either the perirhinal or postrhinal cortices themselves that were originally included in hippocampal formation inputs, but not in cortical inputs by Furtak and colleagues (2007). Similarly we recalculated the relative strength of cortical inputs to the entorhinal cortex to include inputs from both the perirhinal and postrhinal inputs from the data shown in Kerr and colleagues (2007). For figure 45.5, we performed a similar recalculation to show the relative proportion of all cortical (including perirhinal and postrhinal) and subcortical inputs to the rat LEA and MEA from the data shown in Kerr and colleagues (2007).

Abbreviations

35	Area 35 of the perirhinal cortex
36d	Dorsal division of area 36 of the perirhinal cortex
36r	Rostral division of area 36 of the perirhinal cortex
36c	Caudal division of area 36 of the perirhinal cortex
E_C	Caudal division of the entorhinal cortex
E_{CL}	Caudal limiting division of the entorhinal cortex
E_I	Intermediate division of the entorhinal cortex
E_L	Lateral division of the entorhinal cortex
E_O	Olfactory division of the entorhinal cortex
A	Anterior
ER	Entorhinal cortex
IB	Intermediate band
L	Lateral
LB	Lateral band
LEA	Lateral entorhinal area
M	Medial
MB	Medial band
MEA	Medial entorhinal area
Motor	Motor regions of the frontal cortex
OBF	Orbitofrontal cortex
ORB	Orbitofrontal cortex
P	Posterior
PaS	Parasubiculum
PH	Parahippocampal cortex
PIR	Piriform cortex
POR	Postrhinal cortex
PR	Perirhinal cortex
rs	Rhinal sulcus
RSP	Retrosplenial cortex
SS	Somatosensory cortical areas
STG	Superior temporal gyrus
STSd	Dorsal bank of the superior temporal sulcus
STSv	Ventral bank of the superior temporal sulcus
TE/TEO	Visual areas in the ventral temporal lobe of the monkey
TEv	Auditory, somatosensory and visual processing area in the rat temporal lobe
TH	Subdivision of the parahippocampal cortex
TFl	Lateral subdivision of area TF of the parahippocampal cortex
TFm	Medial subdivision of area TF of the parahippocampal cortex
Vis	Visual processing areas in the occipital lobe in rats

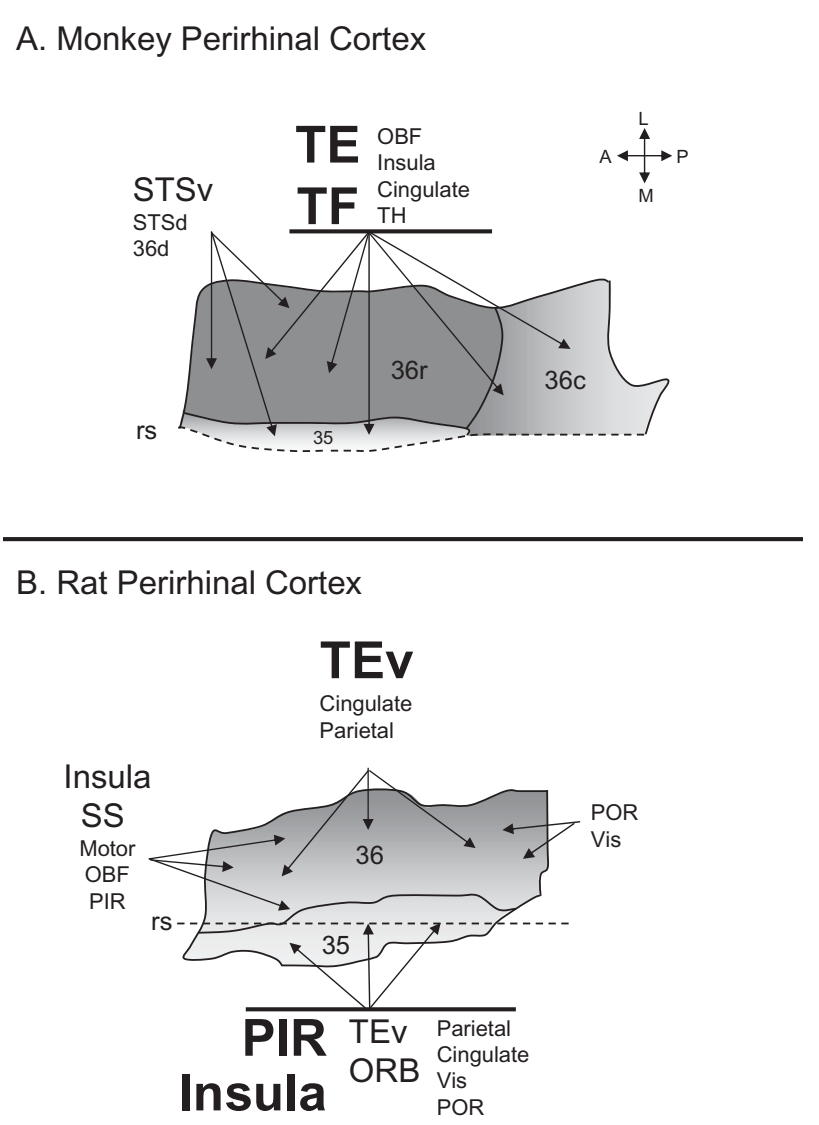


FIGURE 45.2 (*A*) Schematic representation of the topography and strength of cortical inputs to the monkey perirhinal areas 35, 36r, and 36c. The relative strength of the cortical projections is indicated by the size of the lettering, and the locations of the arrows indicate the relative topography of projections throughout the perirhinal cortex. The pattern of intrinsic projections is illustrated by the shading pattern within areas 35, 36r, and 36c. The strongest inputs arise from visual area TE and the adjacent area TF of the parahippocampal cortex that project to all levels of the perirhinal cortex. As indicated by the shading pattern, area 36r projects most strongly to itself and moderately to 36c and 35 and visa versa. (*B*) Schematic representation of the topography and strength of cortical inputs to the rat perirhinal cortex. All conventions are the same as in Panel *A*. Area 36 in rats receives its strongest input from area TEv (mainly from areas processing auditory and somatosensory information) with moderate inputs from insula and somatosensory areas (SS). Area 35 receives its strongest inputs from the piriform cortex (PIR) and insular cortex with moderate inputs from area TEv and the orbitofrontal areas of the frontal lobe (ORB). While dorsal/lateral regions of area 36, project mainly to more medial/ventral regions of area 36, the medial regions of area 36 project strongly to area 35 (shading). These intrinsic projections, however, are not strongly reciprocal. Additional abbreviations: Motor, motor areas of the frontal lobe; POR, postrhinal cortex; rs, rhinal sulcus; SS, somatosensory areas; STG, superior temporal gyrus; STSd, dorsal bank of the superior temporal sulcus; STSv, ventral bank of the superior temporal sulcus.

this convergent cortical input is further processed throughout large extents of the perirhinal cortex.

The Rat Perirhinal Cortex

Cortical afferents In contrast to the preponderance of visual object input to the monkey perirhinal cortex, the rat perirhinal cortex is characterized by a strong convergence of inputs from all sensory modalities (Furtak, Wei, Agster, & Burwell, 2007; Burwell & Amaral, 1998a; Deacon et al., 1983). The strongest inputs to area 36 of the perirhinal cortex arise from anterior and ventral temporal association areas known to receive strong projections from somatosensory (anterior TEv) and auditory areas (mid-rostrocaudal levels of TEv), respectively (Burwell & Amaral, 1998a). The projections from area TEv along with weak projections from cingulate and parietal cortex terminate throughout area 36. Weak projections from the postrhinal cortex as well as weak projections from posterior visual areas both terminate most strongly in caudal portions of area 36. Moderate projections are seen from insular cortex and somatosensory cortical areas (Burwell, 2001; Remple, Henry, & Catania, 2003; Shi & Cassell, 1998) that terminate more strongly in rostral portions of area 36. Frontal areas, including both orbitofrontal areas and frontal motor regions together with the piriform cortex, provide weak projections mainly to anterior levels of area 36. In contrast to area 36, area 35 receives its strongest cortical inputs from piriform cortex and insular cortex, with moderate inputs from TEv and orbitofrontal cortex and weak projections from parietal cortex, cingulate cortex, posterior visual areas, and postrhinal cortex. Thus, taken together, the afferent inputs to rat perirhinal areas 35 and 36 are dominated by sensory inputs from the olfactory, somatosensory, and auditory, as well as the visual, modalities (see box 45.1).

Intrinsic projections While the cortical inputs to area 36 of the rat perirhinal cortex tend to exhibit a more prominent rostrocaudal topography, the intrinsic projections of the perirhinal cortex have a clear dorsal-to-ventral gradient (shaded pattern in figure 45.2*B*; Burwell & Amaral, 1998b). Thus the most dorsal regions of area 36 project mainly to more ventral areas of 36 while the ventral areas of area 36 project strongly to area 35. In contrast, area 35 returns a weaker projection to area 36. Thus area 35 may be the ultimate site of convergence for all sensory modalities within the rat perirhinal cortex.

Summary and comparisons The monkey perirhinal cortex is dominated by high-level visual inputs from area TE as well as prominent polymodal inputs from the parahippocampal cortex. The rat perirhinal cortex, by contrast, receives a much more diverse range of sensory inputs from olfactory, somatosensory, and auditory, as well as visual, modalities. Thus, while monkey perirhinal cortex appears to be specialized in processing visual object information in memory as well as polymodal input from the parahippocampal cortex, the rat perirhinal cortex appears to be poised to integrate information from all sensory modalities in memory.

The Monkey Parahippocampal Cortex

Cortical afferents Like the perirhinal cortex, area TF of the parahippocampal cortex also receives its strongest single input from visual areas (Blatt et al., 2003; Suzuki & Amaral, 1994a), but the visual areas that project to the parahippocampal cortex (mainly areas V4 and TEO; figure 45.3*A*) are posterior to the visual areas that project to the perirhinal cortex (i.e., area TE; figure 45.2*A*). Moreover, these projections exhibit a clear medial lateral topography projecting more strongly to lateral portions of area TF than to medial portions. The next most prominent input to area TF comes from brain areas involved in the so-called ventral visual processing pathway important for analyzing spatial information (the "where" pathway of Ungerleider & Mishkin, 1982). The most prominent of these dorsal stream inputs arise from the retrosplenial cortex, which projects to all levels of area TF. Similarly, moderate projections from area STSd, also considered a dorsal stream area, provide a moderate projection to all levels of area TF. Weak projections are also seen from the posterior parietal cortex, which terminate laterally in area TFl. Moderate projections also originate from area 36c of the perirhinal cortex and weak projections from area 36r that terminate most strongly in anterior portions of area TF while weak projections from area 36d tend to terminate more medially in area TF. Weak projections are also seen from frontal areas including area 46, orbital and medial prefrontal areas (Kondo et al., 2005), and the insular cortex. Area TH exhibits some differences in its cortical inputs relative to the cortical inputs of area TF. The most striking difference is that area TH receives only sparse input from visual area V4, but similar to area TF, it receives prominent projections from the retrosplenial cortex. Another striking difference is the moderate input from auditory association areas of the STG, which appears to constitute the strongest direct auditory projections to any parahippocampal region. Moderate inputs are seen from STSd, and weak inputs arise from insular cortex, the perirhinal cortex (including area 36d), and similar portions of the frontal lobe that project to area TF (see box 45.1).

Intrinsic projection The intrinsic projections of the parahippocampal cortex (illustrated by the shading in figure 45.3*A*) parallel the medial lateral topography of the inputs to this region. Thus area TFl (lateral portions of area TF) has the strongest connections with itself, moderate interconnections with area TFm, and only weak projections with area TH. Similarly, area TFm projects most strongly with itself, but

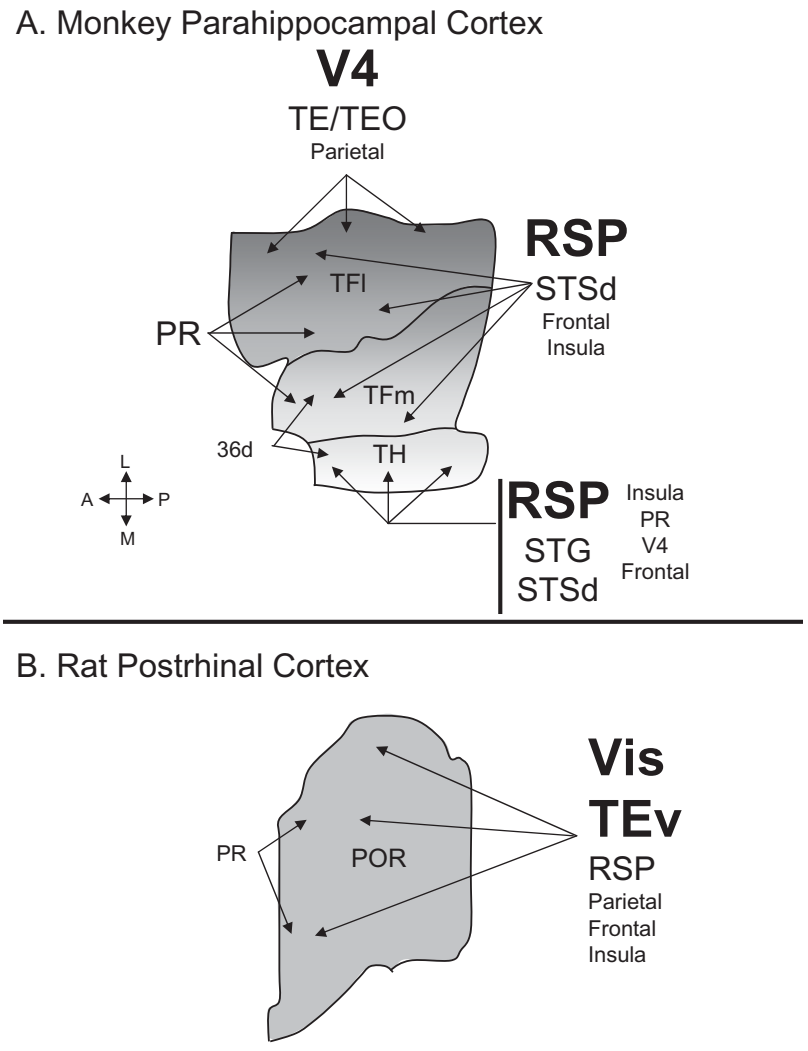


FIGURE 45.3 (*A*) Cortical inputs of the monkey parahippocampal cortical areas TFl, TFm, and TH. All conventions are the same as in figure 45.2. The strongest inputs to area TF arise from visual areas V4 and the retrosplenial cortex (RSP) with moderate inputs from visual areas TE/TEO and polymodal inputs from the dorsal bank of the superior temporal sulcus (STSd). Area TH receives its strongest single input from the retrosplenial cortex (RSP), with moderate input from the superior temporal gyrus (STG) and the dorsal bank of the STS (STSd). Intrinsic projections are strongest within each subdivision and progressively weaker to more distantly located subdivisions (shading). (*B*) The cortical inputs to the rat postrhinal cortex are strongest from primary posterior visual areas (Vis) and area TEv, with moderate inputs from the retrosplenial cortex. No topography of inputs or intrinsic projections is seen. Additional abbreviations: PIR, piriform cortex; SS, somatosensory areas.

moderately with both area TH and area TFl. Finally, like area TFl, area TH projects most strongly with itself, moderately with area TFm, and only weakly with area TFl. Thus, while information arriving only in lateral area TF does not have strong direct interactions with area TH (and vice versa), this information can reach area TH by way of intermediate connections with area TFm.

The Rat Postrhinal Cortex

Cortical afferents Similar to the monkey parahippocampal cortex, the rat postrhinal cortex is dominated by secondary visual inputs from both occipital cortex and temporal lobe visual areas (TEv) as well as from dorsal stream areas including the retrosplenial cortex and parietal cortex. Weak projections are seen from frontal and insular cortices. The perirhinal cortex provides a weak projection to anterior regions of the postrhinal cortex, with the strongest projections arising from area 36. No strong topography of projections to the POR cortex was seen, with all afferent regions projecting to most or all of the postrhinal cortex with the exception of the perirhinal cortex, which tended to project to more anterior regions of the postrhinal cortex. Thus, like the parahippocampal cortex in monkeys, the postrhinal

cortex in rats appears to be a strong site of convergence for both visual and visuospatial input (see box 45.1).

Intrinsic projections The postrhinal cortex does not exhibit any strong topography or polarity in its intrinsic projections. Thus all regions of the postrhinal cortex appear to project to all other regions.

Summary and comparison Striking similarities are seen in the patterns of connections of the monkey parahippocampal cortex and the rat postrhinal cortex. Both areas are dominated by visual input from posterior visual areas together with visuospatial input from so-called dorsal stream areas. More specifically, area TF of the parahippocampal cortex has the strongest resemblance to the postrhinal cortex in rats. In contrast, the two regions differ in that the monkey parahippocampal cortex exhibits a striking topography of inputs and intrinsic connections that is not seen in the rat postrhinal cortex.

The Monkey Entorhinal Cortex

Afferents The cortical inputs of the monkey entorhinal cortex were first studied using anterograde degeneration techniques (Van Hoesen et al., 1975; Van Hoesen & Pandya, 1975a, 1975b) and later using WGA-HRP and fluorescent retrograde tracers (Insausti et al., 1987a). These comprehensive studies showed that the entorhinal cortex is the recipient of prominent input from higher-level polymodal association. If one does not include the temporal pole as part of the perirhinal cortex, then the perirhinal and parahippocampal cortices together make up about half of all the cortical input to the monkey entorhinal cortex, with somewhat more than half of that proportion arising from the parahippocampal cortex and the remainder arising from the perirhinal cortex (Table 1 of Insausti et al., 1987a). The cortex of the temporal pole contributes only weak projections to the entorhinal cortex. Moreover, the perirhinal and parahippocampal cortices exhibit a clear topography of projection, with the parahippocampal cortex projecting most prominently to the posterior entorhinal cortex with weaker projections to anterior and lateral regions and the perirhinal cortex projecting most prominently to anterior and lateral portions of the entorhinal cortex with weaker projections more laterally and caudally. Another prominent input comes from the retrosplenial cortex, which, like the parahippocampal cortex, projects most prominently posteriorly in the entorhinal cortex. Weaker projections are seen from the superior temporal gyrus (STG) and STSd, which project posteriorly, and the insular cortex, which projects anteriorly and laterally. Weak projections are also seen from the piriform cortex to the olfactory subdivision of the entorhinal cortex (E_O). Weak inputs from visual area TE have also been described projecting anteriorly and laterally in the entorhinal cortex (Mohedano-Moriano et al., 2008, 2007). Insausti and colleagues (Mohedano-Moriano et al., 2007) have highlighted the lateral band of the monkey entorhinal cortex as receiving the strongest convergence of afferent input from widespread cortical areas (see box 45.1).

Intrinsic projections and connections with the hippocampus Given that the lateral half of the entorhinal cortex receives the strongest convergent input, another important question concerns how that convergent information is processed intrinsically within the entorhinal cortex. Chroback and Amaral (2007) showed that the intrinsic entorhinal connections in the monkey are organized into rostrocaudally oriented bands where each band extends for about half the anterior-posterior extent of the entorhinal cortex (shaded regions in figure 45.4*A*). There is also a clear medial lateral topography such that two adjacent bands situated end to end cover the lateral portion of the entorhinal cortex, two more bands situated end to end cover the mid-mediolateral portion of the entorhinal cortex, and a single band covers the most rostral and medial entorhinal cortex at the level of area E_O. Interestingly, the projections from the perirhinal and parahippocampal cortices terminate in multiple bands spanning the mediolateral extent of the entorhinal cortex. Additional studies in the monkeys showed that the three mediolaterally oriented bands in the entorhinal cortex project in a topographic fashion to different anterior-posterior levels of the hippocampus as illustrated in figure 45.4*A* (Witter, Van Hoesen, & Amaral, 1989). Thus information originating from the perirhinal and parahippocampal cortices is ultimately processed in the posterior half of the hippocampus.

The Rat Entorhinal Cortex

Afferents The rat entorhinal cortex is subdivided into the medial entorhinal area (MEA) and the lateral entorhinal area (LEA), and these two areas have been further separated into lateral, intermediate, and medial bands that project to different septotemporal regions of the dentate gyrus (Dolorfo & Amaral, 1998b). Given this striking and well-described topography, we will summarize the projections to the entorhinal cortex with respect to the different entorhinal subdivisions (MEA and LEA) as well as the different hippocampal projection bands. By far the strongest cortical input to the LEA originates from the piriform cortex, with moderate inputs also arising from the perirhinal cortex, insula, and frontal cortices. Both the piriform and perirhinal cortices project most strongly to the lateral and intermediate bands and more weakly to the medial band. Parietal cortex and the postrhinal cortex provide a weak input to the LEA with the same overall termination pattern as the piriform and perirhinal cortices. The insular input projects most strongly

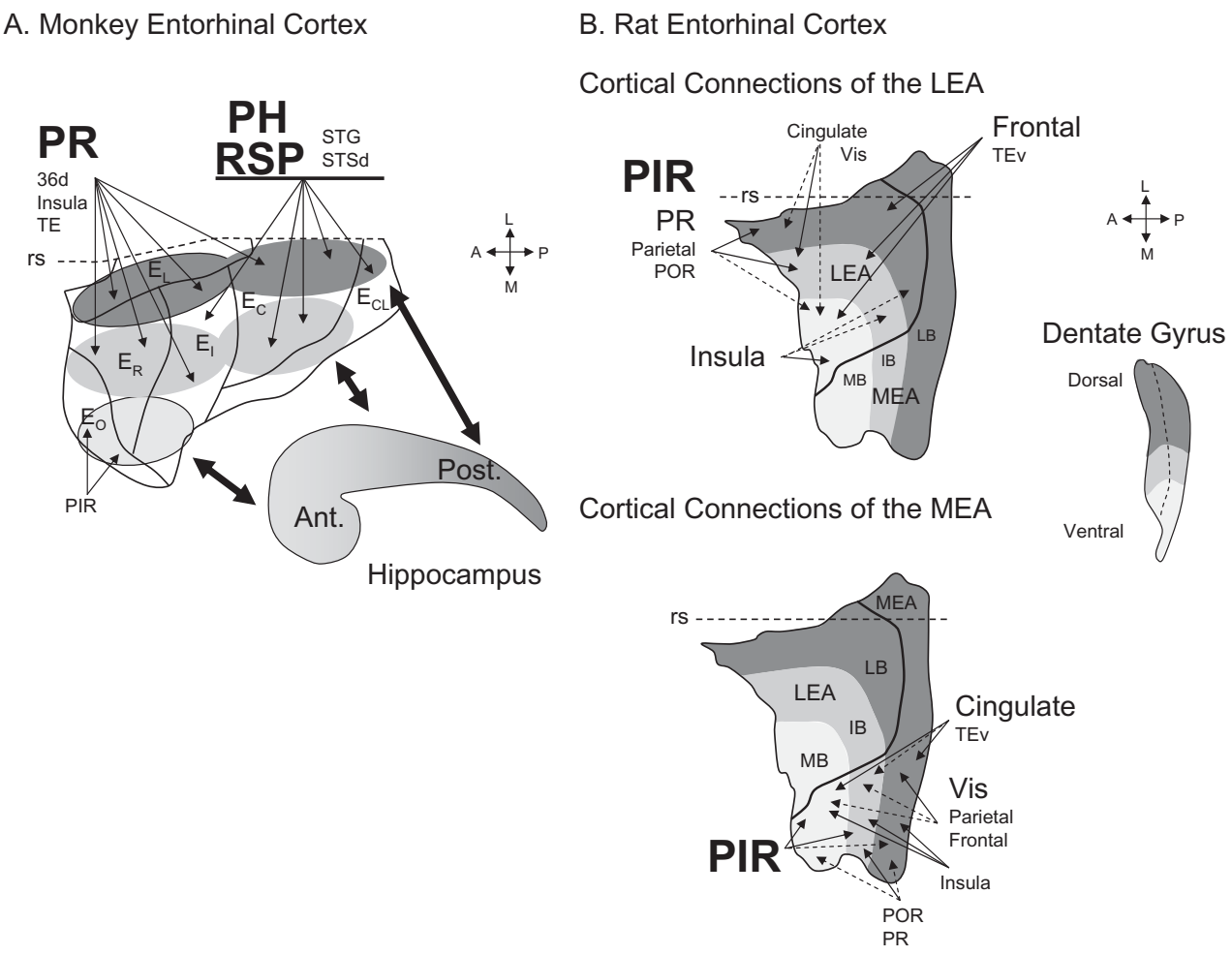


FIGURE 45.4 (*A*) The cortical inputs of the monkey entorhinal cortex arise predominantly from the parahippocampal (PH) and retrosplenial (RSP) cortices that project posteriorly and laterally as well as from the perirhinal cortex (PR) that projects anteriorly and laterally. The shading illustrates three mediolaterally differentiated bands of intrinsic entorhinal projections. The most lateral and middle mediolateral bands (dark and medial dark shades) also exhibit intrinsic connections that are arranged rostrocaudally. The most medial band (lightest shading) has only a single module. These mediolaterally oriented bands also provide a clear topographic projection to different rostro-caudal levels of the hippocampus such that posterior hippocampus receives inputs from the most lateral bands of the entorhinal cortex, mid anterior-posterior hippocampal areas receive projections from the mid-mediolateral bands, and the most anterior portions of the hippocampus receive inputs from the most medially situated entorhinal band. (*B*) *Top:* Illustration of the cortical connections of the lateral entorhinal area (LEA) of the rat. Also illustrated are the locations of the lateral (LB), intermediate (IB), and medial band (MB) that represent both the pattern of intrinsic projections that are maintained mainly within a single band as well as the topographic projection to different dorsoventral levels of the dentate gyrus (shown schematically at the right). The LEA region receives its strongest cortical input from piriform cortex, with moderate projections from the perirhinal cortex, insula, and frontal cortices. The arrows in this figure illustrate the relative strength of projections to different bands with relatively stronger projections illustrated as solid lines and relatively weaker projections illustrated with dashed lines. *Bottom:* Illustration of the cortical inputs to the medial entorhinal area (MEA). All conventions are the same as in the top panel. The strongest cortical input to the MEA originates in the piriform cortex (PIR), with moderate projections seen from the cingulate cortex and posterior visual areas (Vis). Note the weak projections from both the perirhinal (PR) and postrhinal (POR) cortices to the MEA. All additional abbreviations are the same as in figures 45.2 and 45.3.

to the medial band, while the frontal and temporal projections terminate similarly in all three bands. Weak projections from cingulate cortex and visual cortex (Vis) project mainly to the intermediate band.

Similar to the LEA, the most prominent cortical projection of the MEA originates in the piriform cortex, terminating mainly in the medial and intermediate bands. Moderate inputs that arise from the cingulate cortex and visual cortex (Vis) mainly target either the lateral and medial or the lateral band, respectively. Weak inputs are seen from ventral temporal area TEv, and parietal, frontal, and insula cortices. Postrhinal and perirhinal cortices provide only a weak input to the MEA, mainly to the intermediate band (see box 45.1).

Intrinsic connections and connections with the hippocampus Intrinsic projections of the entorhinal cortex tend to remain within one of the roughly rostrocaudally oriented bands of cortex that include both the medial and lateral E_C (i.e., lateral, intermediate, or medial bands—LM, IM, or MB, illustrated in figure 45.4*B*). This finding suggests the possibility that there may be preferential processing of information destined for particular dorsoventral levels of the hippocampus (Kerr, Agster, Furtak, & Burwell, 2007; Dolorfo & Amaral, 1998a). However, like the monkey, the inputs to the perirhinal cortex often cross bands, suggesting that different combinations of information may be processed in different bands and projected to the hippocampus.

Summary and comparison Perhaps the most striking differences in connectivity between the monkey and rat parahippocampal regions are observed at the level of the entorhinal cortex. While the cortical inputs to the monkey entorhinal cortex are dominated by the strong projections from the parahippocampal and perirhinal cortices, the rat LEA receives only moderate inputs from the perirhinal cortex and weak projections from the postrhinal cortex, whereas the MEA receives weak projections from both the perirhinal and postrhinal cortices. Instead, the rat entorhinal cortex is dominated by inputs from olfactory-related areas as well as inputs from the insula, frontal, and cingulate cortices.

Comment on subcortical inputs Another striking difference to consider when evaluating the functional organization of the entorhinal cortex across species is the relative proportion of cortical inputs relative to subcortical inputs. For example, quantitative data from Kerr and colleagues (2007) showed that a defining feature of the rat LEA is that only about one-half of its total afferent inputs (defined as inputs from cortical and nonhippocampal subcortical structures) originate in the cortical areas shown in figure 45.4*B* while the remaining half of its afferent projections arise from subcortical regions including olfactory areas, the claustrum, amygdala, and dorsal thalamus. The relative proportion of subcortical inputs to the rat entorhinal cortex is illustrated in figure 45.5. Similarly, only about half of all afferent inputs to the rat MEA arise from cortical areas. Although a parallel quantification of all cortical and subcortical inputs to the monkey entorhinal cortex has not been done (Insausti et al., 1987a, 1987b), estimations based on the illustrations of Insausti and colleagues (1987b) suggest that subcortical projections make up a much smaller proportion of inputs to the monkey entorhinal cortex compared to the rat entorhinal cortex (figure 45.5). Thus the rat entorhinal cortex not only receives different patterns of cortical inputs but also appears to be influenced much more by its subcortical projections compared to the monkey entorhinal cortex.

Discussion

This comparison of the patterns of cortical inputs, intrinsic projections, and interconnections of the entorhinal, perirhinal/parahippocampal, and postrhinal cortices across both monkeys and rats has revealed both similarities and some striking differences across species (figure 45.5). For example, the perirhinal cortex in both monkeys and rats is characterized by receiving prominent unimodal sensory input. In monkeys this unimodal input is dominated by the visual modality, likely reflecting the predominant role of vision in the primate brain. In contrast, the rat perirhinal cortex receives a much wider range of inputs from all sensory modalities, likely reflecting a species-specific difference in sensory processing. Despite the differences in the range of sensory modalities projecting to the perirhinal cortex in monkeys and rats, damage including the perirhinal cortex in both species is associated with significant deficits in both recognition memory (Mumby, Piterkin, Lecluse, & Lehmann, 2007; Barker, Bird, Alexander, & Warburton, 2007; Mumby & Pinel, 1994; Zola-Morgan, Squire, Clower, & Rempel, 1993; Wood, Mumby, Pinel, & Phillips, 1993; Meunier et al., 1993; Zola-Morgan et al., 1989) and associative memory (Barker & Warburton, 2008; Murray, Gaffan, & Mishkin, 1993; Bunsey & Eichenbaum, 1993). Moreover, perirhinal neurons in both species signal recognition of visual information with a significantly decreased response (Wan, Aggleton, & Brown, 1999; Xiang & Brown, 1998; Brown & Xiang, 1998; Zhu, Brown, & Aggleton, 1995; Miller et al., 1993, 1991; Riches et al., 1991; Baylis & Rolls, 1987; Brown et al., 1987). In addition, it has been well established that monkey perirhinal neurons signal long-term memory for well-learned visual-visual paired associates (Naya, Yoshida, & Miyashita, 2003; Naya, Sakai, & Miyashita, 1996; Sakai & Miyashita, 1991). The patterns of mnemonic activity in the monkey perirhinal cortex suggest the possibility that rat perirhinal neurons may also signal sensory paired associate information as well as recognition memory signals in modalities other than vision (i.e., olfactory, somatosensory, and

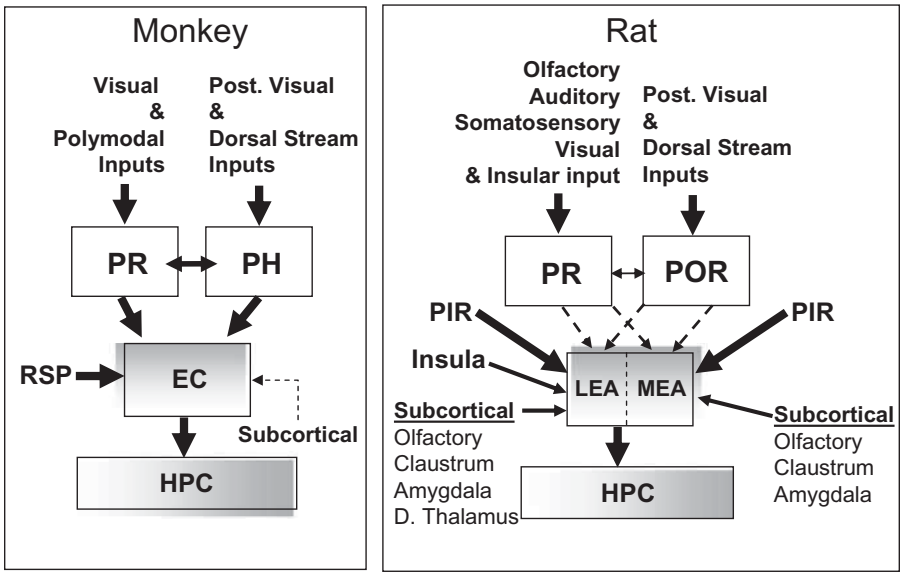


FIGURE 45.5 Schematic illustration of the patterns and relative strength of inputs to the parahippocampal region in monkeys and rats. Similarities are seen in the general patterns of inputs to the perirhinal and parahippocampal/postrhinal cortices in monkeys and rats. However, more striking differences are noted in the patterns of inputs to the entorhinal cortex. To better illustrate one of the key differences, we show the relative strength of the cortical and subcortical projections of the monkey and rat entorhinal cortex, with the relative thickness of the arrows illustrating the relative strength of projections (see box 45.1 for a description of the calculation of the relative strength of cortical inputs). All quantitative data from the rat entorhinal inputs taken from Kerr et al. (2007). However, the weak projections from subcortical regions to the monkey entorhinal cortex are only estimations since similar quantitative comparisons have not been published. Note that the subcortical inputs to the perirhinal (PR) and parahippocampal (PH) or postrhinal (POR) cortices are not illustrated. Additional abbreviations: D. thalamus, dorsal thalamus; RSP, retrosplenial.

auditory). It will be of interest to compare and contrast the full range of mnemonic signals seen in the monkey and rat perirhinal cortex.

Perhaps the most striking cross-species similarities in cortical afferent inputs are seen in the parahippocampal and postrhinal cortices (figure 45.5). These areas in both monkeys and rats receive prominent visual inputs as well as visuospatial inputs from dorsal stream structures including the retrosplenial and parietal cortices. Consistent with these anatomical inputs, the parahippocampal cortex in both monkeys and humans has most commonly been associated with spatial memory functions (Malkova & Mishkin, 2003; Burgess, Maguire, Spears, & O'Keefe, 2001; Bohbot, Allen, & Nadel, 2000; Johnsrude, Owen, Crane, Milner, & Evans, 1999; Maguire, Frackowiak, & Firth, 1997; Aguirre, Detre, Alsop, & D'Esposito, 1996), and in humans it has also been associated with episodic memory (Squire, Stark, & Clark, 2004; Ranganath et al., 2004; Schacter & Wagner, 1999). A growing body of lesion studies in rats suggests a role of the postrhinal cortex in memory for context (Eacott & Easton, 2007; Burwell et al., 2004; Bucci et al., 2002, 2000). However, reports on the contribution of the postrhinal cortex to spatial memory as measured by water maze tasks have been mixed (Burwell et al., 2004; Liu & Bilkey, 2002). Consistent with these lesion studies in rats, the important role of the human parahippocampal cortex in processing contextual associations has recently been highlighted in a series of fMRI studies. Bar and colleagues (Bar, Aminoff, & Ishai, 2008; Aminoff, Gronau, & Bar, 2007; Bar & Aminoff, 2003) report that the parahippocampal cortex is activated in response to objects highly associated with particular contexts (i.e., a traffic light) irrespective of whether the context is spatial (i.e., swings associated with a playground) or nonspatial (i.e., birthday cake associated with a birthday party). Based on these findings, this group has proposed a contextual associative theory of parahippocampal function whereby the parahippocampal cortex is thought to mediate contextual associative processing, an important component of both spatial memory and episodic memory. It will be fascinating to test this contextual associative theory of parahippocampal functions in both monkeys and rats. Neuroanatomical data supports the idea that contextual signals could be seen across both the monkey parahippocampal cortex and the rat postrhinal cortex.

While the parahippocampal/postrhinal cortices in monkeys and rats exhibit clear similarities in their cortical inputs, more striking differences are seen in both the pattern and relative strength of cortical inputs to the entorhinal cortex (figure 45.5). For example, monkey entorhinal inputs are dominated by the prominent projections from the perirhinal and parahippocampal cortices together with strong inputs from the retrosplenial cortex. In contrast, the ento-

rhinal cortex in rats receives its strongest cortical inputs from piriform cortex with moderate inputs from the perirhinal cortex (specifically to LEA), insula, frontal cortex, cingulate cortex, and visual cortical areas. Postrhinal cortex only projects weakly to the entorhinal cortex (figure 45.4*B*). The differences in the patterns of inputs are even more striking if one takes into account that while the monkey entorhinal cortex appears to receive the majority of its inputs from other unimodal and multimodal cortical areas, the rat LEA and MEA receive only about half of all their inputs from cortical structures, with the remaining half arising from subcortical structures (figure 45.5).

Despite the differences in the details of the anatomical connections, there remain intriguing parallels between the rat and monkey entorhinal cortex. For example, it is clear from recent physiological studies focused on the rat entorhinal cortex that cells in LEA and MEA process distinct kinds of information, with the MEA contributing to spatial information, including the striking grid cells specific for the MEA (Hafting et al., 2008; Fyhn et al., 2007; Hafting et al., 2005; Hargreaves, Rao, Lee, & Knierim, 2005; Fyhn et al., 2004). Though recent evidence suggests that cells in LEA are not spatial (Hargreaves et al., 2005), the nature of the input that maximally activates these cells has yet to be identified. The spatial versus nonspatial dissociation in the MEA and LEA in rats parallels the anatomical data in monkeys that the posterior entorhinal cortex receives projections from the strongly visuospatial parahippocampal cortex while the anterior entorhinal cortex receives its strongest projections from the visual-object-processing areas of the perirhinal cortex. These striking topographic projections suggest an anterior-posterior differentiation in object and spatial memory functions, respectively, of the monkey entorhinal cortex (Suzuki & Amaral, 1994b). Indeed, one physiology study attempted to test this anatomical prediction directly by examining responses in the monkey entorhinal cortex during an object version of the delayed-match-to-sample task as well as a spatial version of the same task (Suzuki, Miller, & Desimone, 1997). While neurons throughout the entorhinal cortex signaled memory for both the object and spatial versions of the task, no rostrocaudal topography was found. Because the delayed-match-to-place task required only egocentric and not allocentric spatial memory strategies, this task may not have engaged the particular form of spatial or contextual memory processed by the entorhinal cortex.

The anatomical observations in the monkey entorhinal cortex taken together with the physiological findings from the rat MEA and LEA suggest the possibility that despite the differences in the patterns of inputs, similarities in the functional organization of the entorhinal cortex across species may be present. The entorhinal cortex is one of the least studied medial temporal lobe areas in the monkey, and given the renewed interest in this structure with the discovery of grid cells in the rat MEA, it will be important to continue to explore the functions of the monkey entorhinal cortex. It will also be critical to define any mnemonic role the grid cells may play in the processing of spatial information. Could there be grid cells in the monkey entorhinal cortex, or will they look more like spatial context cells? How will these entorhinal cells in both monkeys and rats participate in memory? Only further studies will tell.

REFERENCES

Aguirre, G. K., Detre, J. A., Alsop, D. C., & D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cereb. Cortex*, *6*, 823–829.

Amaral, D. G., Insausti, R., & Cowan, W. M. (1987). The entorhinal cortex of the monkey. I. Cytoarchitectonic organization. *J. Comp. Neurol.*, *264*, 326–355.

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cereb. Cortex*, *17*, 1493–1503.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*, 347–358.

Bar, M., Aminoff, E., & Ishai, A. (2008). Famous faces activate contextual associations in the parahippocampal cortex. *Cereb. Cortex*, *18*, 1233–1238.

Barker, G. R., Bird, F., Alexander, V., & Warburton, E. C. (2007). Recognition memory for objects, place, and temporal order: A disconnection analysis of the role of the medial prefrontal cortex and perirhinal cortex. *J. Neurosci.*, *27*, 2948–2957.

Barker, G. R., & Warburton, E. C. (2008). NMDA receptor plasticity in the perirhinal and prefrontal cortices is crucial for the acquisition of long-term object-in-place associative memory. *J. Neurosci.*, *28*, 2837–2844.

Baylis, G. C., & Rolls, E. T. (1987). Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. *Exp. Brain Res.*, *65*, 614–622.

Blackstad, T. W. (1956). Commissural connections of the hippocampal region in the rat, with special reference to their mode of termination. *J. Comp. Neurol.*, *105*, 417–537.

Blatt, G. J., Pandya, D. N., & Rosene, D. L. (2003). Parcellation of cortical afferents to three distinct sectors in the parahippocampal gyrus of the rhesus monkey: An anatomical and neurophysiological study. *J. Comp Neurol.*, *466*, 161–179.

Bohbot, V. D., Allen, J. J., & Nadel, L. (2000). Memory deficits characterized by patterns of lesions to the hippocampus and parahippocampal cortex. *Ann. NY Acad. Sci.*, *911*, 355–368.

Brown, M. W., Wilson, F. A. W., & Riches, I. P. (1987). Neuronal evidence that inferomedial temporal cortex is more important than hippocampus in certain processes underlying recognition memory. *Brain Res.*, *409*, 158–162.

Brown, M. W., & Xiang, J. Z. (1998). Recognition memory: Neuronal substrates of the judgement of prior occurrence. *Prog. Neurobiol.*, *55*, 149–189.

Bucci, D. J., Phillips, R. G., & Burwell, R. D. (2000). Contributions of postrhinal and perirhinal cortex to contextual information processing. *Behav. Neurosci.*, *114*, 882–894.

Bucci, D. J., Saddoris, M. P., & Burwell, R. D. (2002). Contextual fear discrimination is impaired by damage to the postrhinal or perirhinal cortex. *Behav. Neurosci.*, *116*, 479–488.

Bunsey, M., & Eichenbaum, H. (1993). Critical role of the parahippocampal region for paired-associate learning in rats. *Behav. Neurosci.*, *107*, 740–747.

Burgess, N., Maguire, E. A., Spiers, H. J., & O'Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *NeuroImage*, *14*, 439–453.

Burwell, R. D. (2000). The parahippocampal region: Corticocortical connectivity. *Ann. NY Acad. Sci.*, *911*, 25–42.

Burwell, R. D. (2001). Borders and cytoarchitecture of the perirhinal and postrhinal cortices in the rat. *J. Comp. Neurol.*, *437*, 17–41.

Burwell, R. D., & Amaral, D. G. (1998a). Cortical afferents of the perirhinal, postrhinal and entorhinal cortices. *J. Comp. Neurol.*, *398*, 179–205.

Burwell, R. D., & Amaral, D. G. (1998b). Perirhinal and postrhinal cortices of the rat: Interconnectivity and connections with the entorhinal cortex. *J. Comp. Neurol.*, *391*, 293–321.

Burwell, R. D., Bucci, D. J., Sanborn, M. R., & Jutras, M. J. (2004). Perirhinal and postrhinal contributions to remote memory for context. *J. Neurosci.*, *24*, 11023–11028.

Burwell, R. D., & Hafeman, D. M. (2003). Positional firing properties of postrhinal cortex neurons. *Neuroscience*, *119*, 577–588.

Burwell, R. D., Saddoris, M. P., Bucci, D. J., & Wiig, K. A. (2004). Corticohippocampal contributions to spatial and contextual learning. *J. Neurosci.*, *24*, 3826–3836.

Burwell, R. D., & Witter, M. P. (2002). Basic anatomy of the parahippocampal region in monkeys and rats. In M. P. Witter & G. Wouterlood (Eds.), *The parahippocampal region: Organization and role in cognitive functions* (pp. 35–60). New York: Oxford University Press.

Burwell, R. D., Witter, M. P., & Amaral, D. G. (1995). The perirhinal and postrhinal cortices of the rat: A review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus*, *5*, 390–408.

Cahusac, P. M., Miyashita, Y., & Rolls, E. T. (1989). Responses of hippocampal formation neurons in the monkey related to delayed spatial response and object-place memory tasks. *Behav. Brain Res.*, *33*, 229–240.

Chrobak, J. J., & Amaral, D. G. (2007). Entorhinal cortex of the monkey. VII. Intrinsic connections. *J. Comp. Neurol.*, *500*, 612–633.

Corkin, S., Amaral, D. G., Gilberto Gonzalez, R., Johnson, K. A., & Hyman, B. T. (1997). H.M.'s medial temporal lobe lesion: Findings from magnetic resonance imaging. *J. Neurosci.*, *17*, 3964–3979.

Deacon, T. W., Eichenbaum, H., Rosenberg, P., & Eckmann, K. W. (1983). Afferent connections of the perirhinal cortex in the rat. *J. Comp. Neurol.*, *220*, 168–190.

Dolorfo, C. L., & Amaral, D. G. (1998a). Entorhinal cortex of the rat: Organization of intrinsic connections. *J. Comp. Neurol.*, *398*, 49–82.

Dolorfo, C. L., & Amaral, D. G. (1998b). Entorhinal cortex of the rat: Topographic organization of the cells to origin of the perforant path projection to the dentate gyrus. *J. Comp. Neurol.*, *398*, 25–48.

Eacott, M. J., & Easton, A. (2007). On familiarity and recall of events by rats. *Hippocampus*, *17*, 890–897.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection*. New York: Oxford University Press.

Furtak, S. C., Wei, S. M., Agster, K. L., & Burwell, R. D. (2007). Functional neuroanatomy of the parahippocampal region in the rat: The perirhinal and postrhinal cortices. *Hippocampus*, *17*, 709–722.

Fyhn, M., Hafting, T., Treves, A., Moser, M. B., & Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, *446*, 190–194.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M. B. (2004). Spatial representation in the entorhinal cortex. *Science*, *305*, 1258–1264.

Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annu. Rev. Psychol.*, *49*, 87–115.

Hafting, T., Fyhn, M., Bonnevie, T., Moser, M. B., & Moser, E. I. (2008). Hippocampus-independent phase precession in entorhinal grid cells. *Nature*, *436*, 801–806.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*, 801–806.

Hargreaves, E. L., Rao, G., Lee, I., & Knierim, J. J. (2005). Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science*, *308*, 1792–1794.

Insausti, R., Amaral, D. G., & Cowan, W. M. (1987a). The entorhinal cortex of the monkey. II. Cortical afferents. *J. Comp. Neurol.*, *264*, 356–395.

Insausti, R., Amaral, D. G., & Cowan, W. M. (1987b). The entorhinal cortex of the monkey. III. Subcortical afferents. *J. Comp. Neurol.*, *264*, 396–408.

Johnsrude, I. S., Owen, A. M., Crane, J., Milner, B., & Evans, A. C. (1999). A cognitive activation study of memory for spatial relationships. *Neuropsychologia*, *37*, 829–841.

Jones, E. G., & Powell, T. P. S. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, *93*, 793–820.

Kerr, K. M., Agster, K. L., Furtak, S. C., & Burwell, R. D. (2007). Functional neuroanatomy of the parahippocampal region: The lateral and medial entorhinal areas. *Hippocampus*, *17*, 697–708.

Kondo, H., Saleem, K. S., & Price, J. L. (2003). Differential connections of the temporal pole with the orbital and medial prefrontal networks in macaque monkeys. *J. Comp. Neurol.*, *465*, 499–523.

Kondo, H., Saleem, K. S., & Price, J. L. (2005). Differential connections of the perirhinal and parahippocampal cortex with the orbital and medial prefrontal networks in macaque monkeys. *J. Comp. Neurol.*, *493*, 479–509.

Krieg, W. J. S. (1946a). Connections of the cerebral cortex. I. The albino rat. A. Topography of the cortical areas. *J. Comp. Neurol.*, *84*, 221–275.

Krieg, W. J. S. (1946b). Connections of the cerebral cortex. I. The albino rat. B. Structure of the cortical areas. *J. Comp. Neurol.*, *84*, 277–323.

Lavenex, P., Suzuki, W. A., & Amaral, D. G. (2004). Perirhinal and parahippocampal cortices of the macaque monkey: Intrinsic projections and interconnections. *J. Comp. Neurol.*, *472*, 371–394.

Leonard, B. W., Amaral, D. G., Squire, L. R., & Zola-Morgan, S. (1995). Transient memory impairment in monkeys with bilateral lesions of the entorhinal cortex. *J. Neurosci.*, *15*, 5637–5659.

Liu, P., & Bilkey, D. K. (2002). The effects of NMDA lesions centered on the postrhinal cortex on spatial memory tasks in the rat. *Behav. Neurosci.*, *116*, 860–873.

Maguire, E. A., Frackowiak, R. S. J., & Firth, C. D. (1997). Recalling routes around London: Activation of the right hippocampus in taxi drivers. *J. Neurosci.*, *17*, 7103–7110.

Malkova, L., & Mishkin, M. (2003). One-trial memory for object-place associations after separate lesions of hippocampus and posterior parahippocampal region in the monkey. *J. Neurosci.*, *23*, 1956–1965.

Martin-Elkins, C. L., & Horel, J. A. (1992). Cortical afferents to behaviorally defined regions of the inferior temporal and parahippocampal gyri as demonstrated by WGA-HRP. *J. Comp. Neurol.*, *321*, 177–192.

Meunier, M., Bachevalier, J., Mishkin, M., & Murray, E. A. (1993). Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *J. Neurosci.*, *13*, 5418–5432.

Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science*, *254*, 1377–1379.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J. Neurosci.*, *13*, 1460–1478.

Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. *Nature*, *273*, 297–298.

Mohedano-Moriano, A., Martinez-Marcos, A., Pro-Sistiaga, P., Blaizot, X., Arroyo-Jimenez, M. M., Marcos, P., Artacho-Perula, E., & Insausti, R. (2008). Convergence of unimodal and polymodal sensory input to the entorhinal cortex in the fascicularis monkey. *Neuroscience*, *151*, 255–271.

Mohedano-Moriano, A., Pro-Sistiaga, P., Arroyo-Jimenez, M. M., Artacho-Perula, E., Insausti, A. M., Marcos, P., Cebada-Sanchez, S., Martinez-Ruiz, J., Munoz, M., Blaizot, X., Martinez-Marcos, A., Amaral, D. G., & Insausti, R. (2007). Topographical and laminar distribution of cortical input to the monkey entorhinal cortex. *J. Anat.*, *211*, 250–260.

Mumby, D. G., & Pinel, J. P. J. (1994). Rhinal cortex lesions and object recognition in rats. *Behav. Neurosci.*, *108*, 11–18.

Mumby, D. G., Piterkin, P., Lecluse, V., & Lehmann, H. (2007). Perirhinal cortex damage and anterograde object-recognition in rats after long retention intervals. *Behav. Brain Res.*, *185*, 82–87.

Murray, E. A., Gaffan, D., & Mishkin, M. (1993). Neural substrates of visual stimulus-stimulus association in rhesus monkeys. *J. Neurosci.*, *13*, 4549–4561.

Murray, E. A., & Mishkin, M. (1986). Visual recognition in monkeys following rhinal cortical ablations combined with either amygdalectomy or hippocampectomy. *J. Neurosci.*, *6*, 1991–2003.

Naya, Y., Sakai, K., & Miyashita, Y. (1996). Activity of primate inferotemporal neurons related to a sought target in pair-association task. *Proc. Natl. Acad. Sci. USA*, *93*, 2664–2669.

Naya, Y., Yoshida, M., & Miyashita, Y. (2003). Forward processing of long-term associative memory in monkey inferotemporal cortex. *J. Neurosci.*, *23*, 2861–2871.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. New York: Oxford University Press.

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*, 2–13.

Remple, M. S., Henry, E. C., & Catania, K. C. (2003). Organization of somatosensory cortex in the laboratory rat (*Rattus norvegicus*): Evidence for two lateral areas joined at the representation of the teeth. *J. Comp. Neurol.*, *467*, 105–118.

Riches, I. P., Wilson, F. A., & Brown, M. W. (1991). The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. *J. Neurosci.*, *11*, 1763–1779.

Rockland, K. S., Saleem, K. S., & Tanaka, K. (1994). Divergent feedback connections from areas V4 and TEO in the macaque. *Visual Neurosci.*, *11*, 579–600.

Rolls, E. T., & Xiang, J. Z. (2005). Reward-spatial view representations and learning in the primate hippocampus. *J. Neurosci.*, *25*, 6167–6174.

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature*, *354*, 152–155.

Saleem, K. S., Kondo, H., & Price, J. L. (2008). Complementary circuits connecting the orbital and medial prefrontal networks with the temporal, insular, and opercular cortex in the macaque monkey. *J. Comp. Neurol.*, *506*, 659–693.

Saleem, K. S., Price, J. L., & Hashikawa, T. (2007). Cytoarchitectonic and chemoarchitectonic subdivisions of the perirhinal and parahippocampal cortices in macaque monkeys. *J. Comp. Neurol.*, *500*, 973–1006.

Saleem, K. S., & Tanaka, K. (1996). Divergent projections from the anterior inferotemporal area TE to the perirhinal and entorhinal cortices in the macaque monkey. *J. Neurosci.*, *16*, 4757–4775.

Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus*, *9*, 7–24.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry*, *20*, 11–21.

Seltzer, B., & Pandya, D. N. (1976). Some cortical projections to the parahippocampal area in the rhesus monkey. *Exp. Neurol.*, *50*, 146–160.

Shapiro, M. L., Tanila, H., & Eichenbaum, H. (1997). Cues that hippocampal place cells encode: Dynamic and hierarchical representation of local and distal stimuli. *Hippocampus*, *7*, 624–642.

Shi, C. J., & Cassell, M. D. (1998). Cascade projections from somatosensory cortex to the rat basolateral amygdala via the parietal insular cortex. *J. Comp. Neurol.*, *399*, 469–491.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.*, *99*, 195–231. [Erratum, *Psychol. Rev.*, *99*(3), 582.]

Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annu. Rev. Psychol.*, *44*, 453–495.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.*, *27*, 279–306.

Stefanacci, L., Buffalo, E. A., Schmolck, H., & Squire, L. R. (2000). Profound amnesia after damage to the medial temporal lobe: A neuroanatomical and neuropsychological profile of patient E.P. *J. Neurosci.*, *20*, 7024–7036.

Suzuki, W. A., & Amaral, D. G. (1994a). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *J. Comp. Neurol.*, *350*, 497–533.

Suzuki, W. A., & Amaral, D. G. (1994b). Topographic organization of the reciprocal connections between monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *J. Neurosci.*, *14*, 1856–1877.

Suzuki, W. A., & Amaral, D. G. (2003a). The perirhinal and parahippocampal cortices of the macaque monkey:

Cytoarchitectonic and chemoarchitectonic organization. *J. Comp. Neurol.*, *463*, 67–91.

Suzuki, W. A., & Amaral, D. G. (2003b). Where are the perirhinal and parahippocampal cortices? A historical overview of the nomenclature and boundaries applied to the primate medial temporal lobe. *Neuroscience*, *120*, 893–906.

Suzuki, W. A., Miller, E. K., & Desimone, R. (1997). Object and place memory in the macaque entorhinal cortex. *J. Neurophys.*, *78*, 1062–1081.

Suzuki, W. A., Zola-Morgan, S., Squire, L. R., & Amaral, D. G. (1993). Lesions of the perirhinal and parahippocampal cortices in the monkey produce long-lasting memory impairment in the visual and tactual modalities. *J. Neurosci.*, *13*, 2430–2451.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In J. Ingle, M. A. Goodale, & R. J. W. Mansfield. (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van Hoesen, G. W., & Pandya, D. N. (1975a). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. III. Efferent connections. *Brain Res.*, *95*, 48–67.

Van Hoesen, G. W., & Pandya, D. N. (1975b). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. I. Temporal lobe afferents. *Brain Res.*, *95*, 1–24.

Van Hoesen, G. W., Pandya, D. N., & Butters, N. (1975). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. II. Frontal lobe afferents. *Brain Res.*, *95*, 25–38.

Wan, H., Aggleton, J. P., & Brown, M. W. (1999). Different contributions of the hippocampus and perirhinal cortex to recognition memory. *J. Neurosci.*, *19*, 1142–1148.

Witter, M. P., Van Hoesen, G. W., & Amaral, D. G. (1989). Topographic organization of the entorhinal projection to the dentate gyrus of the monkey. *J. Neurosci.*, *9*, 216–228.

Wood, E. R., Mumby, D. G., Pinel, J. P. J., & Phillips, A. G. (1993). Impaired object recognition memory in rats following ischemia-induced damage to the hippocampus. *Behav. Neurosci.*, *107*, 51–62.

Xiang, J. Z., & Brown, M. W. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology*, *37*, 657–676.

Zhu, X. O., Brown, M. W., & Aggleton, J. P. (1995). Neuronal signaling of information important to visual recognition memory in rat rhinal and neighboring cortices. *Eur. J. Neurosci.*, *7*, 753–765.

Zola-Morgan, S., Squire, L. R., & Amaral, D. G. (1986). Human amnesia and the medial temporal region: Enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *J. Neurosci.*, *6*, 2950–2967.

Zola-Morgan, S., Squire, L. R., Amaral, D. G., & Suzuki, W. A. (1989). Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. *J. Neurosci.*, *9*, 4355–4370.

Zola-Morgan, S., Squire, L. R., Clower, R. P., & Rempel, N. L. (1993). Damage to the perirhinal cortex exacerbates memory impairment following lesions to the hippocampal formation. *J. Neurosci.*, *13*, 251–265.

46 Medial Temporal Lobe Function and Human Memory

YAEL SHRAGER AND LARRY R. SQUIRE

ABSTRACT The hippocampus and anatomically related structures in the medial temporal lobe support the capacity for conscious recollection (declarative memory). This chapter considers a number of topics that have been prominent in recent discussions of medial temporal lobe function: visual perception, working memory, habit learning, recollection and familiarity, path integration, remote memory, and conscious awareness.

The importance of the medial temporal lobe for memory was established in 1957 when Brenda Milner described the profound effects of medial temporal lobe resection on memory in a patient who became known as HM (Scoville & Milner, 1957; Squire, 2009). Subsequently, animal models of human memory impairment identified the anatomical structures within the medial temporal lobe that are important for understanding HM's memory impairment: the hippocampal region (hippocampus proper, dentate gyrus, and subicular complex) and the perirhinal, entorhinal, and parahippocampal cortices. These structures comprise the medial temporal lobe memory system (Lavenex & Amaral, 2000; Squire & Zola-Morgan, 1991) (figure 46.1*A*).

Medial temporal lobe damage impairs only declarative memory (Schacter & Tulving, 1994; Squire, 1992). Declarative memory refers to the capacity to recollect facts and events. Its contents are thought to be accessible to conscious recollection. The stored representations are flexible and can guide successful performance in a wide range of conditions. Declarative memory can be contrasted with nondeclarative memory, a collection of memory abilities including skills and habits, simple forms of conditioning, priming, and other instances where experience changes how we interact with the world. Nondeclarative memory occurs as modifications within specialized performance systems, and what is learned is expressed through performance rather than recollection. The different forms of nondeclarative memory are supported by specific brain systems outside of the medial temporal lobe memory system (Eichenbaum & Cohen, 2001) (figure 46.1*B*).

YAEL SHRAGER Department of Neurosciences, University of California San Diego, La Jolla, California. Now at Department of Psychology, Harvard University and Howard Hughes Medical Institute

LARRY R. SQUIRE Veterans Affairs Healthcare System, San Diego, California; Department of Psychiatry, Department of Neurosciences, Department of Psychology, University of California, San Diego, La Jolla, California

Intact visual perception

Memory-impaired patients with medial temporal lobe damage have consistently exhibited intact intellectual and perceptual functions. Thus the ability to acquire new memories appears to be a distinct cerebral function, independent of other perceptual and cognitive functions. This fundamental principle of brain organization has been revisited recently, as there has been interest in the possibility that medial temporal lobe structures might be involved in visual perception in addition to memory. Initially the focus was on perirhinal cortex. Whereas some experimental studies with monkeys underscored the role of perirhinal cortex in memory and not visual perception (Buffalo et al., 1999; Hampton & Murray, 2002), others have implicated a role for the perirhinal cortex in visual perception (Buckley, Booth, Rolls, & Gaffan, 2001; Buckley & Gaffan, 1998; Bussey & Saksida, 2002; Bussey, Saksida, & Murray, 2003; Murray & Bussey, 1999). Yet it is difficult to test experimental animals for the ability to identify visual stimuli independent of the ability to learn about them, and it has been pointed out that impairments in monkeys that have been attributed to a perceptual deficit could have resulted from impaired learning (Hampton, 2005).

A distinction between perception and learning can be drawn easily in studies of humans because humans can be instructed about the requirements of the task. A number of studies of patients with medial temporal lobe lesions have found intact perceptual abilities (Holdstock, Gutnikov, Gaffan, & Mayes, 2000; Levy, Shrager, & Squire, 2005; Stark & Squire, 2000). Yet some work in humans found that a group of memory-impaired patients with damage reportedly involving either the hippocampus, or the hippocampus plus additional medial temporal lobe structures, were impaired on tests of perceptual abilities that involved difficult-to-discriminate faces, objects, and scenes (Lee, Buckley, et al., 2005; Lee, Bussey, et al., 2005). This newer work, which involved rather complex visual stimuli, raised the

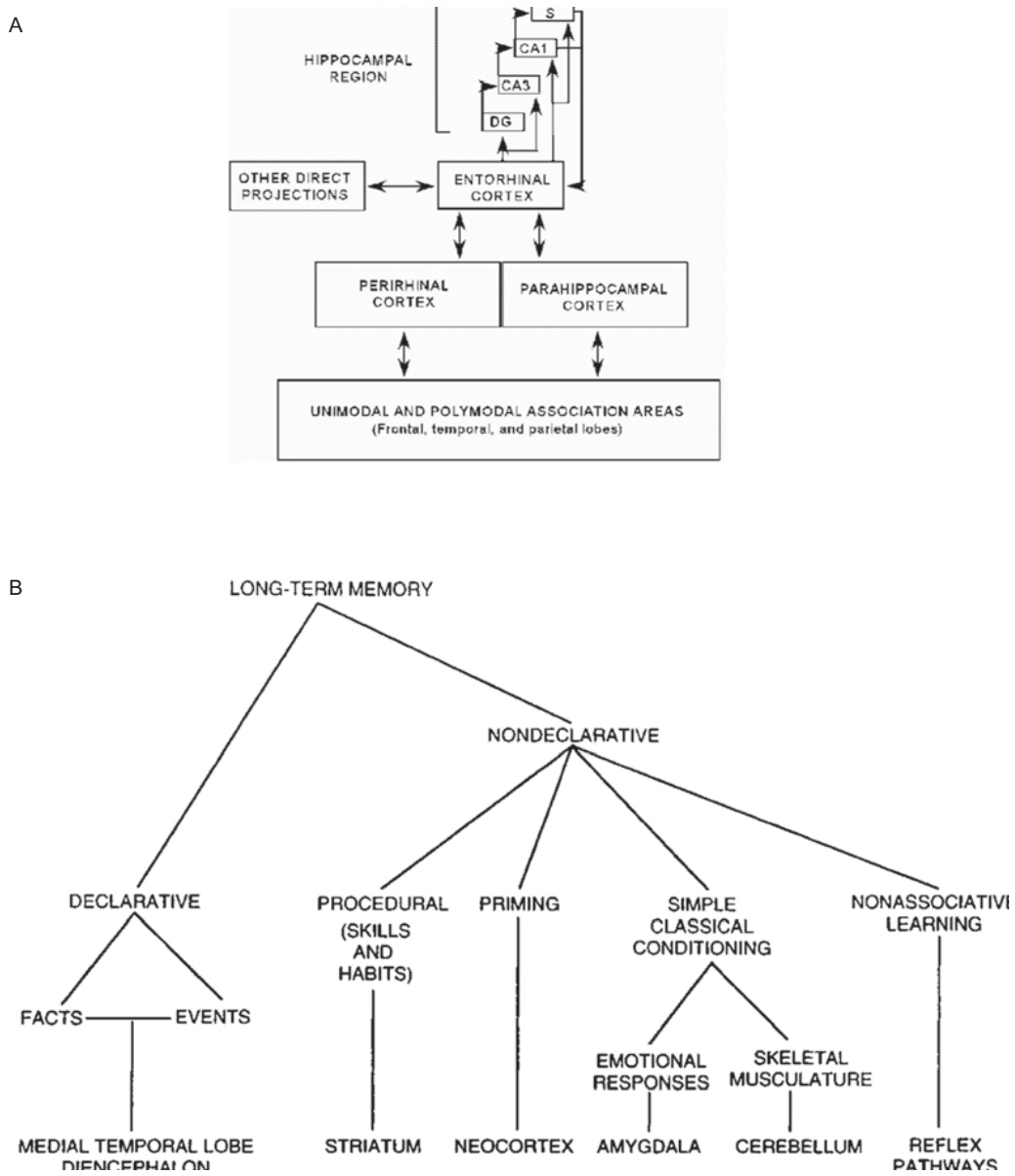


FIGURE 46.1 (*A*) A schematic view of the medial temporal lobe memory system for declarative memory, which is composed of the hippocampal region together with the perirhinal, entorhinal, and parahippocampal cortices. (From Manns & Squire, 2002.) The hippocampal region is composed of the dentate gyrus (DG), the CA fields, and the subiculum (S). (*B*) A taxonomy of mammalian long-term memory systems. The taxonomy lists the brain structures thought to be especially important for each form of declarative and nondeclarative memory. In addition to its central role in emotional learning, the amygdala is able to modulate the strength of both declarative and nondeclarative memory. (From Squire & Knowlton, 2000.)

possibility that appropriate tests can reveal perceptual deficits that had not been detected by conventional tests of visual perception (Lee, Barense, & Graham, 2005). These new findings therefore challenge the long-standing idea that memory impairment can occur as a circumscribed disorder.

Some issues arise in interpreting these studies. First, one wonders if additional damage outside of the medial temporal lobe could underlie the visual perceptual deficits (discussed in the following paragraphs). Second, in these particular studies, the stimuli were created by morphing together two distinct source images, so that the stimuli presented on consecutive trials were derived from the same pair of source images and were therefore quite similar to one another. Accordingly, the question arises whether memory for previous trials could benefit test performance. To test visual perception without testing memory ability, it would be advantageous to use unique stimuli on each trial.

These issues were explored in a recent study of six memory-impaired patients with well-characterized lesions (Shrager, Gold, Hopkins, & Squire, 2006). Two of these

patients (EP and GP) are severely amnesic and have large bilateral lesions of the medial temporal lobe resulting from herpes simplex encephalitis. Both patients have extensive, virtually complete bilateral damage to the hippocampus, amygdala, entorhinal cortex, and perirhinal cortex, as well as the majority of the parahippocampal cortex. Four of the patients have damage thought to be limited to the hippocampus.

The six patients and eight matched controls were tested with morphed grayscale images from three categories (faces, objects, and scenes), similar to those used in the earlier work that reported impairment (Lee, Bussey, et al., 2005). The morphed images were created by gradually morphing one distinct grayscale image into another (e.g., one hat into a different hat or a lemon into a tennis ball) across a 100-step series.

In one experiment, three images were presented on each trial (figure 46.2*A*). Two morphed images were presented below one of the distinct images from which the morphed images were derived, and participants were asked to indicate which of the two morphed images was more similar to the distinct image. Critically, on each trial, each pair of morphed images was derived from a unique pair of distinct images. Thus participants could not benefit from their memory of images they had seen in previous trials. All patients performed as well as controls in all three stimulus categories (faces, objects, and scenes).

On each trial in another experiment, a target image, chosen from the 100-step morphed-image series, was presented at the top of the screen (figure 46.2*B*). In addition, a single image from the same series was presented below the target image. Participants were asked to match the lower image to the target by scrolling though the ordered series of 100 morphed images, viewing only one image at a time, and to select the image that was identical to the target. Performance was scored as the number of image steps between the image that was selected and the target image (thus lower scores indicate better performance). All patients performed as well as controls in all three stimulus categories.

Aside from the possible importance of trial-unique stimuli, it is possible that difference in the patient groups might explain the discrepancy between the findings of Shrager and colleagues (2006) and the findings of Lee and colleagues (Lee, Buckley, et al., 2005; Lee, Bussey, et al., 2005), as well as related findings (Graham et al., 2006). The lesions in the patients studied by Lee and colleagues (Lee, Buckley, et al., 2005; Lee, Bussey, et al., 2005) were characterized by visual ratings of magnetic resonance images (the ratings were made on a 4- or 5-point scale). These ratings, based on visual inspection, are not the same as quantitative brain measurements. Also, the ratings given for each patient were based on a single coronal section for each structure of interest in the medial and lateral temporal lobe, leaving a considerable

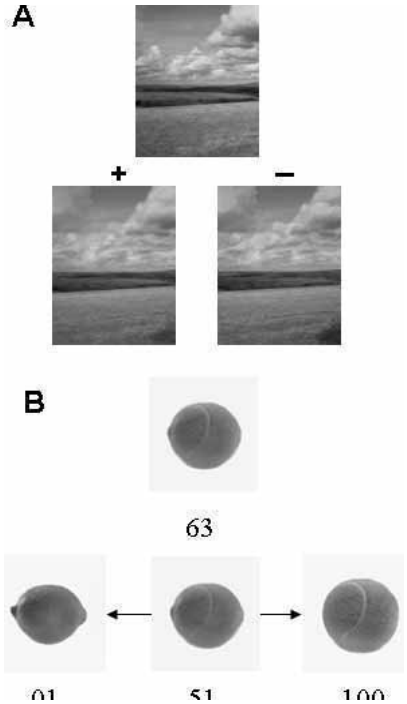


FIGURE 46.2 (*A*) Trial-unique visual discrimination. On each of 120 unique trials, two morphed images were presented below a single distinct image. Participants were asked to choose the lower image (here identified by a +) that appeared more similar to the upper image. (*B*) Visual matching. On each of 45 unique trials, a target image was presented above a single image. Both images were derived from a unique pair of distinct images (01 and 100). In the case illustrated, the target image is image number 63 in the 100-image series, and the bottom image is image number 51 from the same series. Participants were asked to scroll through the ordered series of 100 images to find the image that matched the target image. (From Shrager, Gold, Hopkins, & Squire, 2006.)

amount of tissue unexamined. Furthermore, even by these assessments, the damage in some patients extended beyond the brain structures that defined the groups. Without thorough, quantitative assessment of the lesions, the possibility remains that there is additional damage in the patients and that such damage might underlie the visual perceptual deficits that were observed.

In contrast, the lesions of the patients in Shrager and colleagues (2006) were rigorously measured using quantitative volumetric analysis of magnetic resonance images (Bayley, Gold, Hopkins, & Squire, 2005; Gold & Squire, 2005). For each patient, approximately 60 sections were measured in 1-mm intervals rostrocaudally through the medial and lateral temporal lobes. The measurements were taken in every section in which a structure of interest was present. In addition, volumes were calculated for the insular cortex, the fusiform gyrus, and the frontal, parietal, and occipital lobes.

Over the past 40 years, numerous studies of memory-impaired patients with lesions of the medial temporal lobe

have found visual perceptual function to be intact (Corkin, 1984; Levy et al., 2005; Milner, Corkin, & Teuber, 1968; Stark & Squire, 2000). It was this early work that led to the principle that memory can be severely impaired without impairing other intellectual or perceptual functions. More recently, visual perception has been challenged with newer, more difficult tasks than had been used previously. This new work provides additional support for the principle that memory impairment can occur in the absence of impaired visual perception.

Working memory and brain systems

Working memory refers to the capacity to maintain temporarily a limited amount of information in mind. This information can then be used to support various cognitive abilities, including learning and reasoning (Baddeley & Hitch, 1974). Amnesic patients with damage to the medial temporal lobe have consistently exhibited intact working memory despite grave impairment in long-term memory (Drachman & Arbit, 1966; Milner, 1972). Thus working memory has been thought to be independent of the medial temporal lobe and has come to be defined as a kind of memory that is spared in patients with medial temporal lobe damage (Atkinson & Shiffrin, 1968; Milner, 1972; Pashler & Carrier, 1996).

These ideas have been challenged recently by the proposal that working memory might sometimes depend on medial temporal lobe structures. Specifically, patients with medial temporal lobe damage were found to be impaired at remembering information across brief time intervals (Hannula, Tranel, & Cohen, 2006; Hartley et al., 2007; Nichols, Kao, Verfaellie, & Gabrieli, 2006; Olson, Moore, Stark, & Chatterjee, 2006; Olson, Page, Moore, Chatterjee, & Verfaellie, 2006). The interpretation that these impairments result from impaired working memory would require a revision of the long-standing principle that working memory is separable from long-term memory and is independent of the medial temporal lobe. Yet it is possible that the impairments might have occurred because the capacity for working memory was exceeded and performance in these cases depended on long-term memory. This possibility draws attention to the fact that there is a circularity in the way that working memory is often defined. Working memory has been characterized as a kind of memory that is spared in amnesia, but amnesia is traditionally characterized as a condition in which working memory is intact. It would be useful to have a method for identifying and measuring working memory that is independent of the performance of amnesic patients.

A recent study used distraction between study and test to measure working memory in controls and also tested the performance of amnesic patients (Shrager, Levy, Hopkins, & Squire, 2008). Amnesic patients with medial temporal lobe damage (EP, GP, and six patients with damage limited to the hippocampus) and controls were tested across short delays in four different tasks. Next, the effect of distraction on control performance was tested in the same tasks. The reasoning was as follows: If amnesic patients perform well on tasks when they can operate within working-memory capacity (i.e., by active maintenance), then controls given the same tasks should be impaired when distraction is interposed between study and test because distraction should disrupt the active maintenance process. Conversely, if amnesic patients perform poorly when their working-memory capacity is exceeded, then controls given the same tasks should be minimally affected by distraction between study and test (because performance is now supported more by long-term memory than by active maintenance).

Memory was first tested for names and faces in the patients and their controls. Participants studied either three names presented one at a time or a single face. After a 14-second delay, memory was tested with a single probe stimulus, and participants indicated whether the probe stimulus (a name in the names test and a face in the faces test) had just been presented in the study phase. The patients performed as well as controls in the names test (patients scored 94.4%, and controls scored 94.5% correct), and they were impaired in the faces test (patients scored 93.2%, and controls scored 98.0% correct) (figure 46.3*A*).

The question of interest was whether the impairment in the faces test resulted from a working-memory deficit or a long-term-memory deficit. Accordingly, the effect of distraction on control performance was tested in both the names and faces tests. Controls were again asked to study either three names or a single face. During the delay, on half the trials, controls were distracted. Control performance was impaired by distraction in the names condition (96.4% versus 87.5% correct) but not in the faces condition (96.4% correct for the no-distraction condition versus 95.3% correct for the distraction condition) (figure 46.3*B*). Performance on the distracter (counting) tasks was comparable in the names and faces tests.

These results revealed a correspondence between the performance of amnesic patients and the effect of distraction on controls. Distraction impaired controls on the names test, presumably because the distraction interfered with an active maintenance process that is based on rehearsal. Distraction did not affect performance on the faces test, presumably because the information is difficult to maintain actively (rehearse) and must depend on long-term memory shortly after the information is presented (Warrington & Taylor, 1973). We suggest that amnesic patients were intact when task performance was supported by rehearsal (working memory for names) but were impaired when rehearsal was less effective and performance had to depend on long-term memory (in the case of faces). The same finding was obtained in a related set of experiments that tested memory for objects

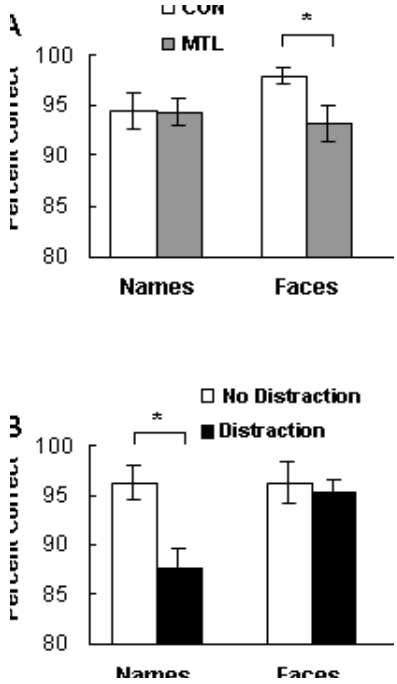


FIGURE 46.3 (*A*) Percent correct scores for controls (CON) and patients with medial temporal lobe lesions (MTL) when asked to remember either three surnames or a single face for 14 seconds. (*B*) Percent correct scores for controls on trials with and without distraction when asked to remember three surnames or a single face for 14 seconds. Error bars indicate standard error. Asterisks indicate $p < 0.05$. (From Shrager, Levy, Hopkins, & Squire, 2008.)

and object locations (a form of relational memory) (Shrager, Levy, et al., 2008). Together, the findings support a brain-based distinction between working memory and long-term memory, as well as the idea that working memory is independent of medial temporal lobe structures.

Habit learning

Some tasks are acquired by humans as declarative knowledge through memorization but nevertheless can be acquired nondeclaratively by experimental animals. On such tasks, amnesic patients with medial temporal lobe lesions perform poorly, whereas monkeys with medial temporal lobe lesions acquire the task at the rate of unoperated monkeys. These findings raise the question whether patients with profound amnesia, with no capacity for declarative memory, could acquire such a task nondeclaratively in the way that the monkey learns it. If so, is the learning done consciously or unconsciously? Or is it the case that, in humans, one memory system cannot readily substitute for another? Perhaps in humans, some forms of nondeclarative memory are not well developed, or perhaps a capacity for nondeclarative learning is easily overridden by the tendency to engage a conscious declarative memory strategy.

One example of a task that is approached differently by humans and nonhuman primates is concurrent discrimination learning, a standard task for studying mammalian memory for more than 50 years. In a common version of this task, eight pairs of objects are presented five times each day, one pair at a time in a mixed order, totaling 40 trials each day. One object in each pair is always correct, and a choice of the correct object results in a reward. Humans readily learn this task after one or two days of training, scoring about 90% correct. The task ordinarily depends on declarative memory, as indicated by the fact that task performance is correlated with the ability to describe the objects and by the fact that amnesic patients perform quite poorly (Hood, Postle, & Corkin, 1999; Squire, Zola-Morgan, & Chen, 1988).

In contrast to the findings in humans, monkeys learned the same concurrent discrimination task gradually, across several hundred trials. Furthermore, monkeys with medial temporal lobe lesions learned this task and a similar version of the same task at normal rates (Buffalo, Stefanacci, Squire, & Zola, 1998; Malamut, Sanders, & Mishkin, 1984; Teng, Stefanacci, Squire, & Zola, 2000). For monkeys, learning proceeded by trial and error (sometimes termed habit learning), and learning was impaired by basal ganglia lesions (Fernandez-Ruiz, Wang, Aigner, & Mishkin, 2001; Teng et al., 2000). Habit memory is proposed to involve slowly acquired associations between stimuli and responses that develop outside of awareness and that are rigidly organized, with the result that what is learned is not readily expressed unless the task is presented just as it was during training.

A recent study asked whether severely amnesic patients can learn this task and, if so, whether the learning has the characteristics of nondeclarative (unconscious) memory. Two patients with large medial temporal lobe lesions, EP and GP, and four controls successfully learned the concurrent discrimination task (eight pairs, five presentations of each pair per session) (Bayley, Frascino, & Squire, 2005). Two sessions were scheduled each week. The controls learned the task quickly during three testing sessions (figure 46.4*A*). In contrast, EP and GP learned gradually during 36 and 28 sessions, respectively, and reached a performance level of 85.0% and 92.5% correct (figure 46.4*B,C*).

The learning exhibited by the patients across weeks was not accompanied by declarative knowledge of the task. Thus neither patient recognized that he had been tested in previous sessions, and neither patient could describe the testing procedure. An additional condition tested whether the knowledge that had been acquired was rigidly organized, as is thought to occur for habit learning, or whether it could be used flexibly. Three to six days after the conclusion of formal training, participants were given a sorting task. They were presented with all 16 objects mixed together (the eight pairs they had learned) and were asked to sort them into two

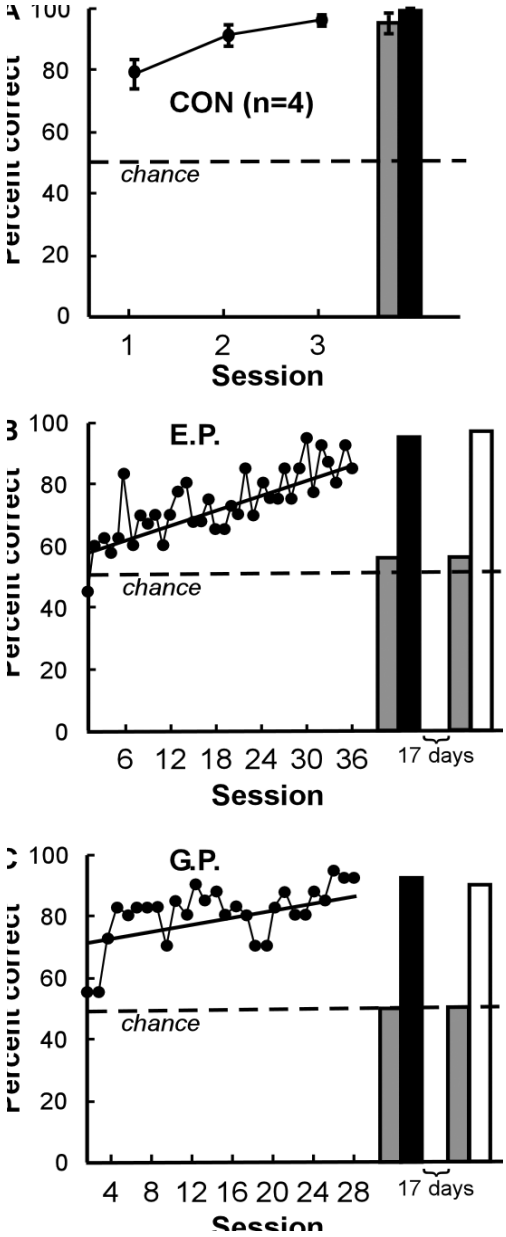


FIGURE 46.4 Performance on the concurrent discrimination task. (*A*) Controls learned the task easily within three sessions and performed well on the sorting task 3–6 days later (gray bar). The black bar shows performance immediately afterward when participants were asked to verbalize their choices rather than reach for objects. Results are means plus and minus standard error. (*B,C*) EP (*B*) and GP (*C*) gradually learned the object pairs across 14–18 weeks. Five days later, each patient failed the sorting task (gray bars) but then, immediately afterward, performed well in the standard task format while verbalizing his responses (black bars). Seventeen days later, both patients again failed the sorting task (gray bars) but performed above 90% when the test was given exactly as in original training (white bars). (From Bayley, Frascino, & Squire, 2005.)

groups: one containing the correct objects and another containing the incorrect objects. Controls succeeded, scoring 95.3% correct, while the patients failed altogether (EP scored 56.3%, and GP scored 50.0% correct; chance = 50%) (figure 46.4*B,C*). Thus the learned information could not be used flexibly. Both patients were able to perform well when asked to verbalize their responses instead of reaching for the objects, but the objects needed to be presented as pairs in order for performance to succeed (figure 46.4*B,C*). Seventeen days later, EP and GP failed the sorting task again and then succeeded once more when the task was presented in its original format (figure 46.4*B,C*).

These findings demonstrated a robust capacity for habit learning that can operate outside awareness and independently of declarative memory and the medial temporal lobe. The knowledge acquired by both patients was rigidly organized and most accessible when the task was structured just as it was during training. These results provide a particularly compelling example of the distinction between declarative (and conscious) and nondeclarative (and unconscious) learning systems.

Recollection and familiarity

One of the most widely studied examples of declarative memory is recognition memory, the capacity to judge an item as having been encountered previously. Recognition memory is thought to consist of two component processes, recollection and familiarity (Mandler, 1980). Recollection involves remembering specific details about the episode in which an item was encountered, whereas familiarity involves simply knowing that an item was presented without any recollection of the original episode.

There has been considerable interest in finding an anatomical basis of the distinction between recollection and familiarity. One suggestion is that the hippocampus is especially important for recollection, whereas the adjacent medial temporal lobe cortex supports familiarity (Brown & Aggleton, 2001; Eichenbaum, Yonelinas, & Ranganath, 2007; Fortin, Wright, & Eichenbaum, 2004; Yonelinas et al., 2002). In contrast, it has also been suggested that the hippocampus is important for both recollection and familiarity (Manns, Hopkins, Reed, Kitchener, & Squire, 2003; Squire, Wixted, & Clark, 2007; Wixted & Squire, 2004).

A recent series of experiments applied signal detection techniques to address these anatomical questions about recollection and familiarity. The receiver operating characteristic (ROC) is a plot of the hit rate versus the false alarm rate across different decision criteria. In order to obtain pairs of hit and false alarm rates at different decision criteria, one can ask participants to provide confidence ratings for their yes/no recognition decisions. A pair of hit and false alarm rates is computed for each level of confidence, and the paired

values are plotted across the confidence levels to construct an ROC.

The ROC of normal individuals has been compared to the ROC of memory-impaired patients (Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998; Yonelinas et al., 2002) and rats with hippocampal lesions (for rats, decision criteria are manipulated by other methods) (Fortin et al., 2004). These ROCs were curvilinear, as is typical, but they differed in their degree of symmetry. As is usually the case, the ROC of controls was asymmetrical, but the ROC of patients and rats with hippocampal lesions was symmetrical (figure 46.5). These data have sometimes been interpreted according to a high-threshold/signal detection model (Yonelinas et al., 1998), which takes the degree of asymmetry in an ROC to reflect the degree to which the recollection process contributes to recognition memory performance. Specifically, a symmetrical ROC indicates that recollection was absent and that recognition memory was based only on familiarity, whereas an asymmetrical ROC indicates that recollection also occurred to some extent. Thus, by the high-threshold/signal detection model, the finding that memory-impaired patients, as well as rats with hippocampal lesions, produce a symmetrical ROC suggests that the recollection process is impaired.

Although the ROC curves of patients and their controls (and lesioned rats and their controls) did differ qualitatively with respect to symmetry, they also differed quantitatively. The patients and the lesioned rats had weaker memories than their respective controls. Indeed, the standard signal detection model of recognition memory (Macmillan & Creelman, 2005) explains the difference between asymmetrical and symmetrical ROCs as a difference in memory strength. An asymmetrical ROC reflects high memory strength, and a symmetrical ROC reflects lower memory strength (Glanzer, Kim, Hilford, & Adams, 1999). If the symmetry of the ROC is related to memory strength, then the difference in symmetry between controls and memory-impaired patients (or lesioned rats) might simply reflect the difference between strong and weak memories, rather than a qualitative difference between the underlying component processes of recognition memory.

This idea was tested in a study of controls and memory-impaired patients with circumscribed hippocampal lesions (Wais, Wixted, Hopkins, & Squire, 2006). The question of interest was how the shape of the ROC changes as a function of memory strength for patients with hippocampal lesions and how the performance of patients compares with the performance of controls. If recollection is selectively impaired in the patients, then the ROC should be symmetrical regardless of memory strength. Alternatively, if the hippocampus does not selectively support recollection, then the patients with hippocampal lesions should produce asymmetrical ROCs like the controls once differences in memory strength are accounted for.

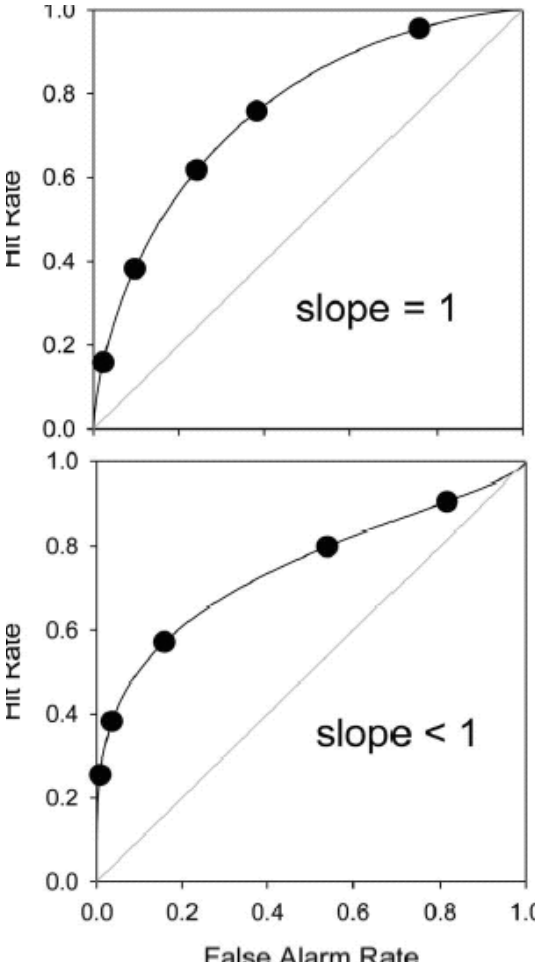


Figure 46.5 Hypothetical ROC data illustrating symmetrical and asymmetrical ROC curves. The degree of asymmetry evident in an ROC is typically quantified by a "slope" parameter obtained by fitting the standard signal detection model (Macmillan & Creelman, 2005) to the data. A slope of 1.0 denotes a symmetrical ROC, whereas a slope less than 1.0 denotes an asymmetrical ROC. The high-threshold/signal detection model (Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998) would yield a recollection parameter estimate of 0 for the symmetrical ROC (*top panel*) and an estimate greater than 0 for the asymmetrical ROC (*bottom panel*). (From Wais, Wixted, Hopkins, & Squire, 2006.)

Six patients with damage thought to be limited to the hippocampus participated. Participants first studied 50 words. After a three-minute interval, 50 target words were intermixed with 50 foil words, and participants assigned a confidence rating to each word from 1 ("definitely new") to 6 ("definitely old"). As expected, the patients performed more poorly than controls (H-50 versus C-50, figure 46.6). Patients were then given a second, easier recognition-memory test involving only 10 words (plus four untested filler words, two at the beginning and two at the end of the list). On this test, patient performance improved to a level similar to that of controls (H-10 versus C-50, figure 46.6).

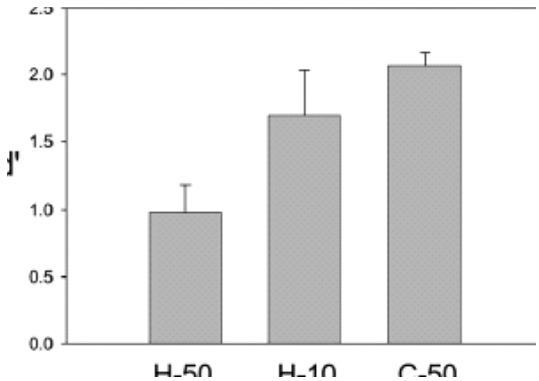


FIGURE 46.6 Recognition memory performance of hippocampal patients and controls. Patients were tested with 50-item lists (H-50 condition) or 10-item lists (H-10 condition). Controls were tested with 50-item lists (C-50 condition). The retention interval was 3 minutes. The mean score of the controls (C-50) was greater than that of the patients in the H-50 condition, but similar to the score obtained by the patients in the H-10 condition. The score in the H-10 condition was also greater than the score in the H-50 condition. Error bars represent standard errors. (From Wais, Wixted, Hopkins, & Squire, 2006.)

The ROCs for the patients and controls were all curvilinear (figure 46.7). The ROC from the H-50 condition was symmetrical, but the ROCs from the H-10 and the C-50 conditions were asymmetrical to a similar extent. Thus the ROC of the hippocampal patients was symmetric when memory was weak but asymmetric when memory was strong (H-50 versus H-10, respectively). Moreover, when memory performance was similar for patients and controls (the H-10 and C-50 conditions), the degree of asymmetry in the ROC was similar as well.

To derive theoretical estimates of recollection and familiarity, the ROC data were first fitted by the high-threshold/signal detection model. In the H-50 condition, the recollection parameter estimate was equal to zero, and in the C-50 condition it was greater than zero (0.23). Similarly, the familiarity parameter estimate was lower in the H-50 condition than in the C-50 condition (0.83 versus 1.64). Importantly, in the H-10 condition, the parameter estimates for both recollection and familiarity were similar to the estimates for the C-50 condition (recollection estimate of 0.22 and 0.23 for H-10 and C-50, respectively, and familiarity estimate of 1.21 and 1.64 for H-10 and C-50, respectively, $p = 0.11$). Thus, according to the high-threshold/signal detection model, the recollection process is present in both patients and controls. Furthermore, when memory performance was matched between patients and controls (H-10 and C-50), the nearly identical recollection estimates (0.22 and 0.23) offered no evidence of a selective deficit in recollection after hippocampal lesions.

In contrast to the high-threshold/signal detection model, the traditional signal detection model (Macmillan &

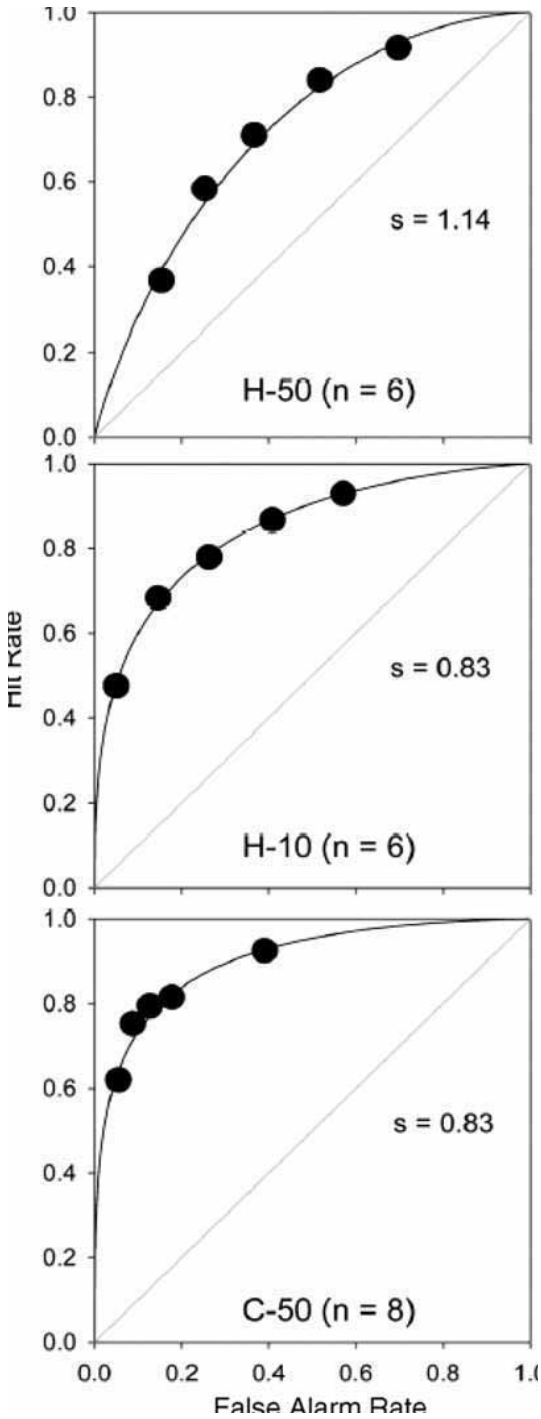


FIGURE 46.7 ROC data produced by the hippocampal patients and controls. The top panel shows the data for hippocampal patients in the 50-item condition, the middle panel shows the data for hippocampal patients in the 10-item condition, and the bottom panel shows the data for controls in the 50-item condition. The H-50 ROC was symmetric (slope = 1.14). The H-10 ROC and the C-50 ROC were both asymmetric (slope = 0.83 for both groups) and also more asymmetric than the ROC of the H-50 group. (From Wais, Wixted, Hopkins, & Squire, 2006.)

Creelman, 2005) does not dictate how recollection and familiarity combine to produce an ROC curve. The fact that patients and controls exhibited similar ROCs as a function of memory strength nevertheless suggests that the component processes of recognition are both operative in the patients. If the asymmetry of the ROC curve is taken as an indicator of recollection, then these results challenge the idea that the hippocampus subserves a recollection process and that hippocampal patients do not have this process. The findings are not an argument against the utility of the constructs of recollection and familiarity. Rather, they challenge the idea that recollection and familiarity can be dichotomized and assigned to separate brain structures in the medial temporal lobe (Squire et al., 2007).

Path integration

During the past several decades, there have been two influential traditions about the function of the hippocampus, entorhinal cortex, and related medial temporal lobe structures. One tradition emphasizes the importance of these structures for memory (Scoville & Milner, 1957; Squire, Stark, & Clark, 2004). The other emphasizes their importance for spatial cognition (Etienne & Jeffery, 2004; McNaughton, Battaglia, Jensen, Moser, & Moser, 2006; O'Keefe & Nadel, 1978; Whitlock, Sutherland, Witter, Moser, & Moser, 2008). An important part of spatial cognition is path integration, the ability to use internal cues during movement (i.e., self-motion cues) to keep track of a reference location. Because many tasks of spatial cognition, including path integration, require memory, these two traditions are compatible with each other to a large extent.

The view that medial temporal lobe structures are important for memory makes a key distinction between short-term (or working) memory and long-term memory (see the section on working memory in this chapter). Patients with damage to the medial temporal lobe, including damage to the hippocampus or entorhinal cortex, are thought to have intact working memory, and they perform poorly only when demands are made on long-term memory. This idea is meant to apply even to tasks that require spatial cognition, such as path integration.

In contrast, the view that the hippocampus and entorhinal cortex are important for path integration often includes the suggestion that the path integrator is located in these structures (Etienne & Jeffery, 2004; McNaughton et al., 2006). By this view, patients with damage to the hippocampus and entorhinal cortex should be impaired at path integration, and this impairment should occur regardless of whether demands are made on long-term memory.

These ideas were tested by asking whether the hippocampus and entorhinal cortex are essential for path integration even when the task can be managed within working memory (Shrager, Kirwan, & Squire, 2008). Two patients with large medial temporal lobe lesions (EP and GP), three patients with hippocampal lesions, and seven controls were tested for their path integration ability. In the first condition (standard), participants wore a blindfold and earphones to reduce external cues, and they were led in a laboratory space along 16 different paths that averaged 4.3 meters in length and involved either 1 or 2 turns. At the end of each path, participants stepped onto a platform (5 cm above the floor and equipped with handlebars for stability) and were asked to point to their start location. An error measure was then computed as the difference between the participant's pointing direction and the correct direction. Participants were encouraged to actively maintain the path in mind as they walked, so that performance might be supported by working memory (mean trial duration was 33.4 seconds).

The patients performed as accurately as controls (mean pointing direction for patients, −4°; controls, +4°) (figure 46.8*A*). Furthermore, the variability in performance across the 16 trials was similar for both groups (patients, 31.3; controls, 30.5) (figure 46.8*D*). Debriefings of the two most severely memory-impaired patients (EP and GP) and four controls indicated that subjects tried to keep track of their position in space as they moved, continually updating their position relative to the start point.

Path integration was further challenged in two additional conditions. In one condition, participants were blindfolded and led in the laboratory along 16 paths involving 3 turns (compared to 1 or 2 turns in the standard condition; mean trial duration, 26.0 seconds). In another condition, participants were blindfolded and led in an outdoor space along 8 paths that were nearly four times as long (15 meters) as the paths in the standard condition. Mean trial duration in this case was 29.7 seconds. In both conditions, the patients pointed to their start location as accurately as controls, and they also exhibited variability similar to that of controls (for 3 turns, controls, +6°, variability, 32.2; patients, −7°, variability, 31.7; for the longer paths, controls, +9°, variability, 35.0; patients, −15°, variability, 27.2). In a fourth condition, participants were led along 8 paths in the laboratory environment (4 involving 1 turn, 4 involving 2 turns, for a path length averaging 4.2 m). At the end of each path, participants estimated their distance from the start location (instead of pointing). Some paths ended far from the start point, and some ended near the start point. Again, the patients were as accurate as controls (both groups averaged 0.7 m error for distances that averaged 2.8 m).

A separate condition served as a key control to ensure that participants were in fact path integrating, that is, relying on internal cues rather than on external cues that were beyond experimental control. Blindfolded participants were led in the laboratory environment along 16 paths and, at the end of each path, stepped onto the platform and held onto the

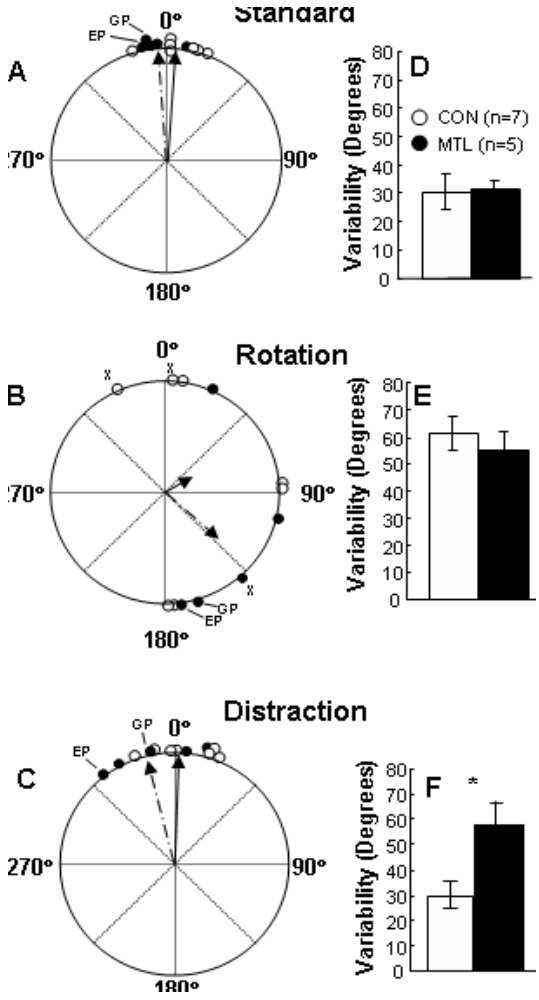


FIGURE 46.8 Circular means of each participant's 16 pointing directions in the standard, rotation, and distraction conditions for patients with damage to the medial temporal lobe (MTL, filled circles) and controls (CON, unfilled circles). 0° indicates the correct direction. Group pointing directions are also indicated (solid arrow, CON; broken arrow, MTL). Shorter arrows denote greater variability (dispersion) in the group's pointing direction (following Moore's test for nonuniformity, Batschelet, 1981). In *B, X* indicates individuals who did not exhibit a significant pointing direction. The standard deviation of pointing directions around each participant's circular mean was calculated, and the individual standard deviations were then averaged for each group (*D,E,F*). Asterisk (*) indicates $p < 0.05$, CON versus MTL groups. Brackets indicate standard error. (From Shrager, Kirwan, & Squire, 2008.)

handlebars. The platform was then slowly rotated by remote control through 190°, after which participants tried to point to their start location. Pilot experiments indicated that, after the rotation, participants had difficulty knowing how far they had been turned. Accordingly, one would expect that, if participants were in fact relying on path integration (internal cues) to point to their start location, they should have difficulty in the rotation condition. Mean trial duration was the same as in the original, standard condition (32.4 seconds). The result was that performance of both patients and controls was substantially compromised. Neither group exhibited a significant pointing direction (i.e., pointing across participants was random), and variability increased (patients, 54.9; controls, 61.5) (figure 46.8*B,E*).

In a final condition, path integration was tested when demands on long-term memory were increased by increasing the duration of each trial and by introducing distraction during the delay (mean trial duration, 1 minute 10 seconds). The controls performed as well in the distraction condition as in the standard condition. Their mean pointing direction was +1° (compared to +4° in the standard condition), and variability was 30.1 (compared to 30.5 in the standard condition) (figure 46.8*C,F*). In contrast, the patients had difficulty in the distraction condition. Their mean pointing direction was −14° (numerically worse than their pointing direction in the standard condition, −4°), and variability was 57.1 (significantly worse than in the standard condition, 31.3, and significantly worse than controls in the distraction condition, 30.1) (figure 46.8*C,F*).

These results indicate that patients with lesions of the medial temporal lobe can path integrate as well as controls when the task can be managed within working memory. When demands on long-term memory were increased, the patients were impaired. These findings suggest that medial temporal lobe structures are not unique, essential sites where computations needed for path integration are carried out. These computations likely occur upstream of the medial temporal lobe, perhaps in parietal cortex. The medial temporal lobe then operates on this information, much as it operates on information from other sensory modalities, in order to transform on-line perceptual information into long-term memory.

Remote memory

Damage to the hippocampus and related medial temporal lobe structures not only impairs new learning capacity but also impairs memory for information acquired before the damage occurred (retrograde amnesia). Early clinical descriptions of retrograde amnesia led to the proposal that recently acquired memories are typically more impaired than remotely acquired memories (Ribot, 1881), and a large experimental literature has supported this idea (Frankland & Bontempi, 2005; Squire & Bayley, 2007). Yet questions remain about whether medial temporal lobe damage can sometimes cause extensive and ungraded retrograde memory loss and about the status of remote autobiographical memory.

Some have concluded that retrograde amnesia is temporally ungraded and that recent and remote memories are similarly impaired across the life span (Sanders & Warrington, 1971; Warrington, 1996). Others have concluded that retrograde amnesia is temporally limited and related to the extent and locus of the damage (Eichenbaum,

Dudchenko, Wood, Shapiro, & Tanila, 1999; Squire et al., 2004). There are two reasons why this issue has been difficult to settle. First, memory has not always been assessed at early enough time periods to permit a firm conclusion that memory loss is ungraded. Second, the relationship between the extent of retrograde memory loss and the extent of medial temporal lobe damage has not always been clearly identified.

A recent study tested memory for past news events in two patients with large medial temporal lobe lesions (EP and GP), six patients with limited hippocampal lesions, and matched controls (Bayley, Hopkins, & Squire, 2006). The test involved up to 300 questions about news events that had occurred from early life to the current year. The patients with hippocampal lesions performed poorly during the period of anterograde amnesia (after the onset of amnesia) and exhibited temporally limited retrograde amnesia covering a period of about 5 years before the onset of amnesia (figure 46.9). For more remote time periods, the patients performed as well as controls.

EP and GP also performed poorly during their period of anterograde amnesia and in addition exhibited extensive retrograde amnesia covering many years before the onset of amnesia (figure 46.10). Nevertheless, both patients performed better when the questions covered the most remote time periods. GP performed within 1.1 standard deviations of controls in the time period 21 to 25 years before amnesia (when he would have been 17 to 21 years old), and he performed as well as controls in the time period 26 to 30 years before amnesia. EP reached normal levels of performance when the questions covered the period 46 to 50 years before amnesia (when he would have been 20 to 24 years old).

With respect to autobiographical memory, it has been proposed (usually in single-case studies) that medial temporal lobe damage, and even limited hippocampal damage, leads to impaired memory for personal events that extends into early life (Cipolotti et al., 2001; Hirano & Noguchi, 1998; Moscovitch, Nadel, Winocur, Gilboa, & Rosenbaum, 2006; Steinvorth, Levine, & Corkin, 2005). Findings from group studies, however, suggest that both patients with limited hippocampal lesions and patients with large medial temporal lobe lesions have intact autobiographical memory of early life (Bayley, Gold, et al., 2005; Bayley, Hopkins, & Squire, 2003; Bright et al., 2006; Eslinger, 1998; Rempel-Clower, Zola, Squire, & Amaral, 1996). For example, a study of six patients with limited hippocampal lesions, two patients with large medial temporal lobe lesions, and 25 controls were given 24 cue words, and for each word were asked to recollect a specific event from the first third of their lives that involved the word (Bayley et al., 2003). Narratives were first scored on a 4-point scale (scores of 0 to 3 points) according to how well participants described an event that was specific to time and place. Patients and controls produced a similar number of well-formed (3-point) memories

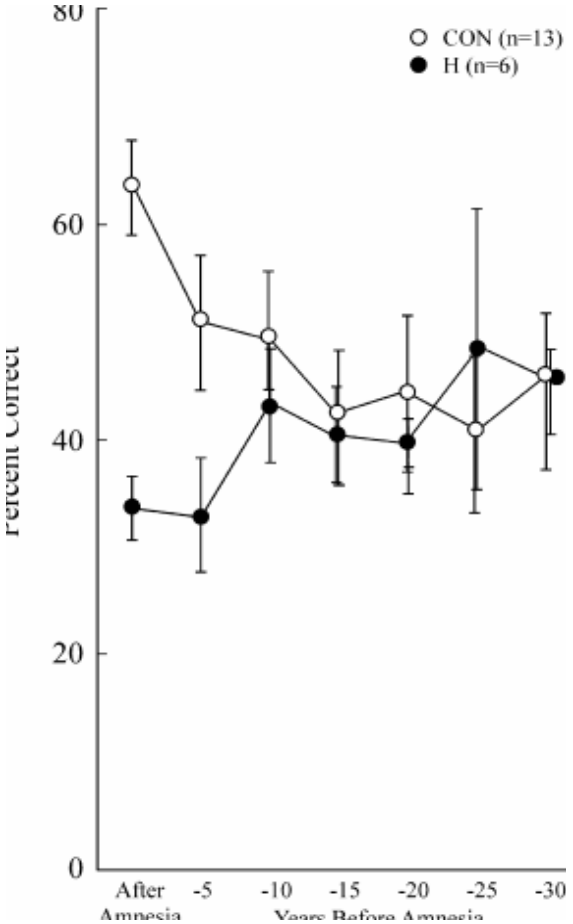


FIGURE 46.9 Recall performance on a test of 279 news events that occurred from 1951 to 2005. The scores for controls (CON) and six patients with damage limited to the hippocampus (H) have been aligned relative to the onset of amnesia so that performance can be shown for the time period after the onset of amnesia and in 5-year intervals for the time preceding the onset of amnesia. The data point at −5 represents 1–5 years before amnesia, the point at −10 represents 6–10 years before amnesia, and so on. Error bars indicate standard error. (From Bayley, Hopkins, & Squire, 2006.)

and were able to produce 3-point memories in response to most of the key words. The narratives were then submitted to a detailed analysis of content. The narratives of patients and controls contained the same number of details and were similar on several other measures as well.

In an effort to maximize the sensitivity with which the assessment of remote memory is carried out, it is also possible to use techniques that ask for a single memory from a given time period (instead of 24 memories, as before) and then probe extensively to obtain as many as 50 details for each memory (the Autobiographical Interview; Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002). This test was given to three patients with damage limited to the hippocampus, two patients with large lesions of the medial temporal lobe, and five controls (Kirwan, Bayley, Galván, & Squire, 2008).

Participants were asked to provide one memory from each of five time periods: childhood (up to age 11 years), teenage years (age 12–17), early adulthood (age 18–35), middle age (age 36–55), and the year before testing. The patients with hippocampal lesions were impaired only at the most recent time period, and the patients with larger medial temporal lobe lesions were impaired at the two most recent time periods (figure 46.11). Both groups of patients performed as well as controls in the three earliest time periods.

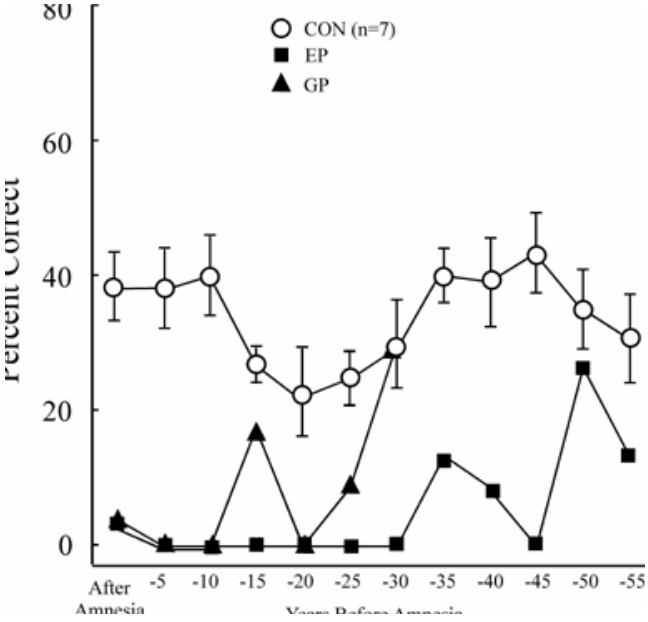


FIGURE 46.10 Recall performance on a test of news events that occurred from 1938 to 2005 (for EP, 300 events) and from 1951 to 2005 (for GP, 279 events). The scores for the two patients with large medial temporal lobe lesions and controls (CON) have been aligned relative to the onset of amnesia (see caption for figure 46.9). Error bars indicate standard error. (From Bayley, Hopkins, & Squire, 2006.)

Impaired remote autobiographical memory does occur when the brain damage extends beyond the medial temporal lobe. Another study assessed remote autobiographical memory in three patients with medial temporal lobe damage plus significant damage to the neocortex (Bayley, Gold, et al., 2005). As in an earlier study (Bayley et al., 2003), the patients were asked to recall a childhood memory in response to a cue word for each of 24 words. As described previously, patients with damage limited to the medial temporal lobe and their controls produced well-formed memories in response to most of the 24 word cues (21.6 for the patients, 22.9 for the controls). In contrast, the patients with significant neocortical damage outside of the medial temporal lobe were severely impaired and provided a mean of only 4.0 unique, well-formed memories. These patients were able to recall some general information in response to the cue words but had marked difficulty providing memories that were specific to a particular time and place.

Similar findings were obtained with the Autobiographical Memory Interview (AMI; Kopelman, Wilson, & Baddeley, 1989), a standardized test that facilitates comparison of performance across laboratories. In the childhood portion of this test, patients are asked to recall three unique events from their childhood. Patients with medial temporal lobe damage plus significant neocortical damage performed poorly, whereas patients with limited medial temporal lobe damage performed well (Bayley, Gold, et al., 2005; figure 46.12). These findings suggest that patients who fail the AMI (Childhood Portion), or who otherwise have difficulty recollecting events from their early life, have damage outside the medial temporal lobe.

Awareness and memory

Declarative memory has ordinarily been viewed as memory that is accompanied by knowledge or awareness of what

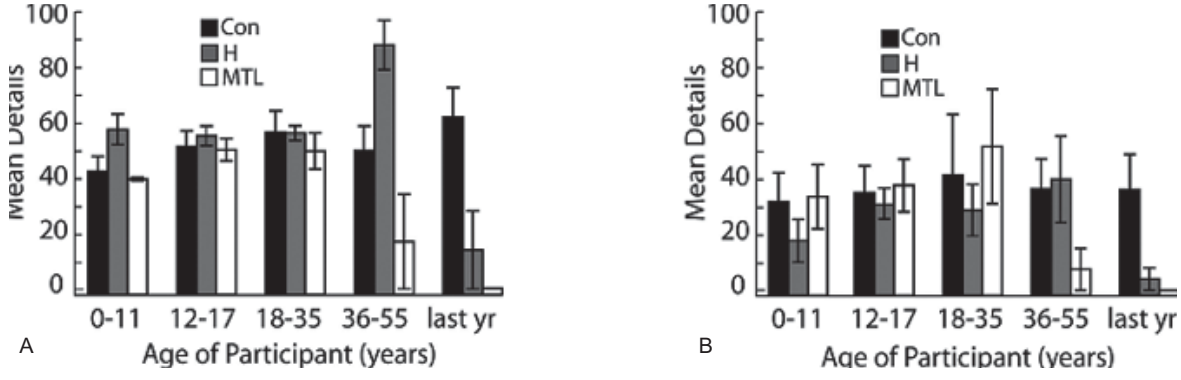


FIGURE 46.11 Total number of (*A*) episodic and (*B*) semantic details across time periods. Patients with damage thought to be limited to the hippocampus (H), patients with larger medial temporal lobe lesions (MTL), and controls (Con) were asked to retrieve one autobiographical memory from each of five time periods. Error bars indicate standard error. (From Kirwan, Bayley, Galván, & Squire, 2008.)

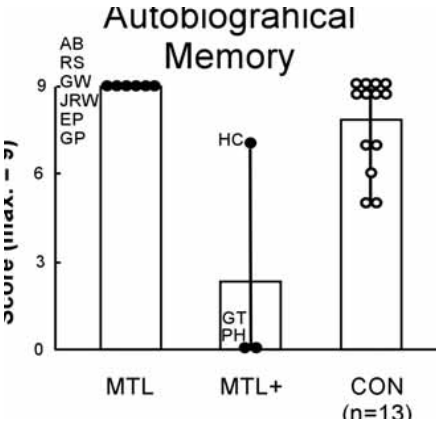


FIGURE 46.12 Performance on the Childhood Portion of the Autobiographical Memory Interview (maximum score, 9). Each participant's score is represented by a circle, and patients are identified by initials. MTL, patients with medial temporal lobe lesions; MTL+, patients with medial temporal lobe lesions and additional lesions in neocortex; CON, controls. (From Bayley, Gold, Hopkins, & Squire, 2005.)

has been learned, and the availability of learned material to awareness has been considered one of its key features (Eichenbaum, 1997; Gabrieli, 1998; Squire, 1992; Tulving & Schacter, 1990). In some cases, when behavior is changed by experience, it is unclear what kind of memory is being expressed. Consider the case of eye movements. When individuals view novel scenes, familiar scenes, or familiar scenes in which a change has been introduced, eye movements differ depending on the viewing history of each scene. What kind of memory is indexed by eye movements, and is this kind of memory accessible to awareness?

Two studies addressed this issue by asking what kind of memory is operating when eye movements change as the result of experience (Smith, Hopkins, & Squire, 2006; Smith & Squire, 2008). Amnesic patients and controls viewed scenes that were either novel, repeated, or manipulated. For manipulated scenes, an element was either added to or removed from a previously presented scene. The first finding was that the patients were impaired at remembering whether a scene was new or old and also whether a scene had been manipulated or not. The second finding was that control subjects, but not amnesic patients, examined scenes differently depending on whether the scenes were new or old. Specifically, during a 5-second viewing period, controls made fewer fixations and sampled fewer regions when scenes were repeated than when they were novel. Importantly, these effects occurred only when individuals were aware that a scene was novel or repeated. The third finding was that, when scenes were manipulated, healthy participants made more fixations in the manipulated region, spent more time looking at the manipulated region, and made more transitions into and out of the manipulated region than in corresponding, unmanipulated regions in the repeated scenes. Again, these effects occurred only when individuals were aware of the manipulation. Participants who were unaware that a scene had been changed looked at it in the same way that they looked at repeated scenes (figure 46.13). The fourth finding was that these effects occurred even when the scenes were presented without any indication that memory was being tested or that individuals should try to detect which scenes were new, old, or manipulated.

Thus there was no indication that eye movements reveal an unaware (unconscious) form of memory. Instead, eye movements reflected declarative (conscious) memory. These findings support the principle that hippocampus-dependent memory is accessible to awareness. Recent studies of transitive inference (Smith & Squire, 2005) and eyeblink conditioning (Smith, Clark, Manns, & Squire, 2005) are consistent with this idea. See Smith, Hopkins, and Squire (2006) for discussion of two studies that reached different conclusions (Greene, Spellman, Dusek, Eichenbaum, & Levy, 2001; Ryan, Althoff, Whitlow, & Cohen, 2000).

Summary

This chapter reviewed a number of recent findings pertinent to the organization of memory and the function of the medial temporal lobe. These findings indicated that (1) visual perception is independent of the medial temporal lobe; (2) working memory can be identified independently of the performance of amnesic patients, and when this identification is accomplished, working memory is found to be independent of the medial temporal lobe; (3) humans have a robust capacity for habit learning that operates outside of awareness and is independent of the medial temporal lobe; (4) path integration, a form of spatial cognition, is independent of the hippocampus and entorhinal cortex when information can be maintained within working memory, but path integration depends on these structures when demands are made on long-term memory; (5) the hippocampus is required for recognition memory, regardless of whether decisions are based on recollection or familiarity; (6) the medial temporal lobe plays a time-limited role in declarative memory such that very remote memory, including remote autobiographical memory, is intact after medial temporal lobe damage; and (7) the kind of declarative memory that is dependent on the hippocampus is accessible as conscious, aware knowledge of what has been learned.

ACKNOWLEDGMENTS This work was supported by the Medical Research Service of the Department of Veterans Affairs, NIMH, the Metropolitan Life Foundation, and an NSF predoctoral fellowship (Y.S.).

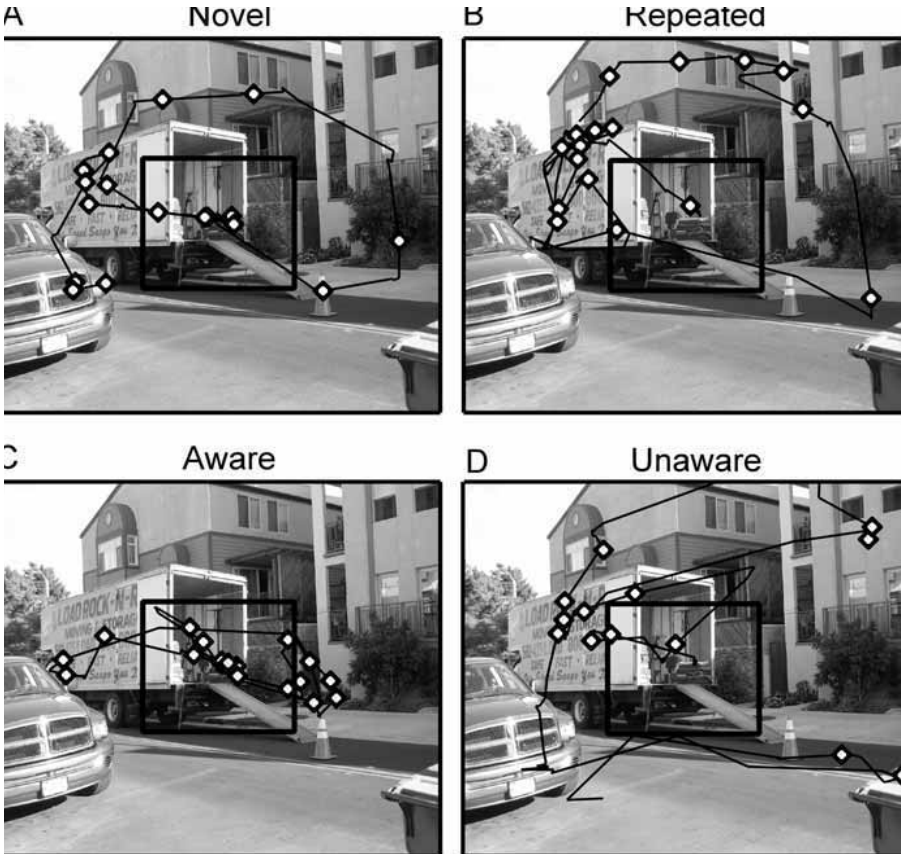


FIGURE 46.13 Eye movement traces (black lines) and fixations (diamonds) for four different individuals who viewed this image for 5 seconds. (*A*) An individual for whom this image was novel. (*B*) An individual for whom this image was familiar. (*C*) An individual for whom the image was familiar but altered (a man with a dolly is no longer in the back of the truck) and who was aware of the change. (*D*) An individual for whom the image was familiar but altered and who was unaware of the change. Aware individuals spent more time looking at the altered region of the image than unaware individuals or individuals who had never seen the version with the man in the truck. In each panel, the critical region is identified by a black square, but the square did not appear during testing. (From Smith, Hopkins, & Squire, 2006.)

REFERENCES

ATKINSON, R. C., & SHIFFRIN, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89–195). New York: Academic Press.

BADDELEY, A., & HITCH, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47–89). New York: Academic Press.

BATSCHELET, E. (1981). *Circular statistics in biology*. London: Academic Press.

BAYLEY, P. J., FRASCINO, J. C., & SQUIRE, L. R. (2005). Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature*, *436*, 550–553.

BAYLEY, P. J., GOLD, J. J., HOPKINS, R. O., & SQUIRE, L. R. (2005). The neuroanatomy of remote memory. *Neuron*, *46*, 799–810.

BAYLEY, P. J., HOPKINS, R. O., & SQUIRE, L. R. (2003). Successful recollection of remote autobiographical memories by amnesic patients with medial temporal lobe lesions. *Neuron*, *38*, 135–144.

BAYLEY, P. J., HOPKINS, R. O., & SQUIRE, L. R. (2006). The fate of old memories after medial temporal lobe damage. *J. Neurosci.*, *26*, 13311–13317.

BRIGHT, P., BUCKMAN, J., FRADERA, A., YOSHIMASU, H., COLCHESTER, A. C., & KOPELMAN, M. D. (2006). Retrograde amnesia in patients with hippocampal, medial temporal, temporal lobe, or frontal pathology. *Learn. Memory*, *13*, 545–557.

BROWN, M. W., & AGGLETON, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nat. Rev. Neurosci.*, *2*, 51–61.

BUCKLEY, M. J., BOOTH, M. C., ROLLS, E. T., & GAFFAN, D. (2001). Selective perceptual impairments after perirhinal cortex ablation. *J. Neurosci.*, *21*, 9824–9836.

BUCKLEY, M. J., & GAFFAN, D. (1998). Perirhinal cortex ablation impairs visual object identification. *J. Neurosci.*, *18*, 2268–2275.

BUFFALO, E. A., RAMUS, S. J., CLARK, R. E., TENG, E., SQUIRE, L. R., & ZOLA, S. M. (1999). Dissociation between the effects of damage to perirhinal cortex and area TE. *Learn. Memory*, *6*, 572–599.

BUFFALO, E. A., STEFANACCI, L., SQUIRE, L. R., & ZOLA, S. M. (1998). A reexamination of the concurrent discrimination learning task: The importance of anterior inferotemporal cortex, area TE. *Behav. Neurosci.*, *112*, 3–14.

BUSSEY, T. J., & SAKSIDA, L. M. (2002). The organization of visual object representations: A connectionist model of effects of lesions in perirhinal cortex. *Eur. J. Neurosci.*, *15*, 355–364.

Bussey, T. J., Saksida, L. M., & Murray, E. A. (2003). Impairments in visual discrimination after perirhinal cortex lesions: Testing "declarative" vs. "perceptual-mnemonic" views of perirhinal cortex function. *Eur. J. Neurosci.*, *17*, 649–660.

Cipolotti, L., Shallice, T., Chan, D., Fox, N., Scahill, R., Harrison, G., Stevens, J., & Rudge, P. (2001). Long-term retrograde amnesia . . . the crucial role of the hippocampus. *Neuropsychologia*, *39*, 151–172.

Corkin, S. (1984). Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in H.M. *Sem. Neurol.*, *4*, 249–258.

Drachman, D. A., & Arbit, J. (1966). Memory and the hippocampal complex. II. Is memory a multiple process? *Arch. Neurol.*, *15*, 52–61.

Eichenbaum, H. (1997). Declarative memory: Insights from cognitive neurobiology. *Annu. Rev. Psychol.*, *48*, 547–572.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain.* New York: Oxford University Press.

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, *23*, 209–226.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.*, *30*, 123–152.

Eslinger, P. (1998). Autobiographical memory after temporal lobe lesions. *Neurocase*, *4*, 481–495.

Etienne, A. S., & Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus*, *14*, 180–192.

Fernandez-Ruiz, J., Wang, J., Aigner, T. G., & Mishkin, M. (2001). Visual habit formation in monkeys with neurotoxic lesions of the ventrocaudal neostriatum. *Proc. Natl. Acad. Sci. USA*, *98*, 4196–4201.

Fortin, N. J., Wright, S. P., & Eichenbaum, H. (2004). Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature*, *431*, 188–191.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nat. Rev. Neurosci.*, *6*, 119–130.

Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annu. Rev. Psychol.*, *49*, 87–115.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. (1999). Slope of the receiver-operating characteristic in recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.*, *25*, 500–513.

Gold, J. J., & Squire, L. R. (2005). Quantifying medial temporal lobe damage in memory-impaired patients. *Hippocampus*, *15*, 79–85.

Graham, K. S., Scahill, V. L., Hornberger, M., Barense, M. D., Lee, A. C., Bussey, T. J., & Saksida, L. M. (2006). Abnormal categorization and perceptual learning in patients with hippocampal damage. *J. Neurosci.*, *26*, 7547–7554.

Greene, A. J., Spellman, B. A., Dusek, J. A., Eichenbaum, H. B., & Levy, W. B. (2001). Relational learning with and without awareness: Transitive inference using nonverbal stimuli in humans. *Mem. Cogn.*, *29*, 893–902.

Hampton, R. R. (2005). Monkey perirhinal cortex is critical for visual memory, but not for visual perception: Reexamination of the behavioural evidence from monkeys. *Q. J. Exp. Psychol. [B]*, *58*, 283–299.

Hampton, R. R., & Murray, E. A. (2002). Learning of discriminations is impaired, but generalization to altered views is intact, in monkeys (*Macaca mulatta*) with perirhinal cortex removal. *Behav. Neurosci.*, *116*, 363–377.

Hannula, D. E., Tranel, D., & Cohen, N. J. (2006). The long and the short of it: Relational memory impairments in amnesia, even at short lags. *J. Neurosci.*, *26*, 8352–8359.

Hartley, T., Bird, C. M., Chan, D., Cipolotti, L., Husain, M., Vargha-Khadem, F., & Burgess, N. (2007). The hippocampus is required for short-term topographical memory in humans. *Hippocampus*, *17*, 34–48.

Hirano, M., & Noguchi, K. (1998). Dissociation between specific personal episodes and other aspects of remote memory in a patient with hippocampal amnesia. *Percept. Mot. Skills*, *87*, 99–107.

Holdstock, J. S., Gutnikov, S. A., Gaffan, D., & Mayes, A. R. (2000). Perceptual and mnemonic matching-to-sample in humans: Contributions of the hippocampus, perirhinal and other medial temporal lobe cortices. *Cortex*, *36*, 301–322.

Hood, K. L., Postle, B. R., & Corkin, S. (1999). An evaluation of the concurrent discrimination task as a measure of habit learning: Performance of amnesic subjects. *Neuropsychologia*, *37*, 1375–1386.

Kirwan, C. B., Bayley, P. J., Galván, V. V., & Squire, L. R. (2008). Detailed recollection of remote autobiographical memory after damage to the medial temporal lobe. *Proc. Natl. Acad. Sci. USA*, *105*, 2676–2680.

Kopelman, M. D., Wilson, B. A., & Baddeley, A. D. (1989). The Autobiographical Memory Interview: A new assessment of autobiographical and personal semantic memory in amnesic patients. *J. Clin. Exp. Neuropsychol.*, *11*, 724–744.

Lavenex, P., & Amaral, D. G. (2000). Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus*, *10*, 420–430.

Lee, A. C., Barense, M. D., & Graham, K. S. (2005). The contribution of the human medial temporal lobe to perception: Bridging the gap between animal and human studies. *Q. J. Exp. Psychol. [B]*, *58*, 300–325.

Lee, A. C., Buckley, M. J., Pegman, S. J., Spiers, H., Scahill, V. L., Gaffan, D., Bussey, T. J., Davies, R. R., Kapur, N., Hodges, J. R., & Graham, K. S. (2005). Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus*, *15*, 782–797.

Lee, A. C., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., Hodges, J. R., & Graham, K. S. (2005). Perceptual deficits in amnesia: Challenging the medial temporal lobe "mnemonic" view. *Neuropsychologia*, *43*, 1–11.

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychol. Aging*, *17*, 677–689.

Levy, D. A., Shrager, Y., & Squire, L. R. (2005). Intact visual discrimination of complex and feature-ambiguous stimuli in the absence of perirhinal cortex. *Learn. Memory*, *12*, 61–66.

Macmillan, N., & Creelman, C. (2005). *Detection theory: A user's guide.* Mahwah, NJ: Lawrence Erlbaum Associates.

Malamut, B. L., Saunders, R. C., & Mishkin, M. (1984). Monkeys with combined amygdalo-hippocampal lesions succeed in object discrimination learning despite 24-hour intertrial intervals. *Behav. Neurosci.*, *98*, 759–769.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychol. Rev.*, *87*, 252–271.

Manns, J. R., Hopkins, R. O., Reed, J. M., Kitchener, E. G., & Squire, L. R. (2003). Recognition memory and the human hippocampus. *Neuron*, *37*, 171–180.

Manns, J. R., & Squire, L. R. (2002). The medial temporal lobe and memory for facts and events. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), *The handbook of memory disorders* (pp. 81–99). New York: John Wiley & Sons.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the "cognitive map." *Nat. Rev. Neurosci.*, *7*, 663–678.

Milner, B. (1972). Disorders of learning and memory after temporal lobe lesions in man. *Clin. Neurosurg.*, *19*, 421–446.

Milner, B., Corkin, S., & Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. *Neuropsychologia*, *6*, 215–234.

Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., & Rosenbaum, R. S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Curr. Opin. Neurobiol.*, *16*, 179–190.

Murray, E. A., & Bussey, T. J. (1999). Perceptual-mnemonic functions of the perirhinal cortex. *Trends Cogn. Sci.*, *3*, 142–151.

Nichols, E. A., Kao, Y. C., Verfaellie, M., & Gabrieli, J. D. (2006). Working memory and long-term memory for faces: Evidence from fMRI and global amnesia for involvement of the medial temporal lobes. *Hippocampus*, *16*, 604–616.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford, UK: Clarendon.

Olson, I. R., Moore, K. S., Stark, M., & Chatterjee, A. (2006). Visual working memory is impaired when the medial temporal lobe is damaged. *J. Cogn. Neurosci.*, *18*, 1087–1097.

Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., & Verfaellie, M. (2006). Working memory for conjunctions relies on the medial temporal lobe. *J. Neurosci.*, *26*, 4596–4601.

Pashler, H., & Carrier, M. (1996). Memory. In E. Bjork & R. Bjork (Eds.), *Handbook of perception and cognition* (pp. 3–29). San Diego: Academic Press.

Rempel-Clower, N. L., Zola, S. M., Squire, L. R., & Amaral, D. G. (1996). Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. *J. Neurosci.*, *16*, 5233–5255.

Ribot, T. (1881). *Les maladies de la memoire* [Diseases of memory]. New York: Appleton-Century-Crofts.

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychol. Sci.*, *11*, 454–461.

Sanders, H. I., & Warrington, E. K. (1971). Memory for remote events in amnesic patients. *Brain*, *94*, 661–668.

Schacter, D. L., & Tulving, E. (1994). *Memory systems*. Cambridge, MA: MIT Press.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry*, *20*, 11–21.

Shrager, Y., Gold, J. J., Hopkins, R. O., & Squire, L. R. (2006). Intact visual perception in memory-impaired patients with medial temporal lobe lesions. *J. Neurosci.*, *26*, 2235–2240.

Shrager, Y., Kirwan, C. B., & Squire, L. R. (2008). The neural basis of the cognitive map: Path integration does not require hippocampus or entorhinal cortex. *Proc. Natl. Acad. Sci. USA, 105*, 12034–12038.

Shrager, Y., Levy, D. A., Hopkins, R. O., & Squire, L. R. (2008). Working memory and the organization of brain systems. *J. Neurosci.*, *28*, 4818–4822.

Smith, C., & Squire, L. R. (2005). Declarative memory, awareness, and transitive inference. *J. Neurosci.*, *25*, 10138–10146.

Smith, C. N., Clark, R. E., Manns, J. R., & Squire, L. R. (2005). Acquisition of differential delay eyeblink classical conditioning is independent of awareness. *Behav. Neurosci.*, *119*, 78–86.

Smith, C. N., Hopkins, R. O., & Squire, L. R. (2006). Experience-dependent eye movements, awareness, and hippocampus-dependent memory. *J. Neurosci.*, *26*, 11304–11312.

Smith, C. N., & Squire, L. R. (2008). Experience-dependent eye movements for old and new scenes reflect hippocampus-dependent, declarative memory, even in the absence of memory instructions. *J. Neurosci., 28*, 12825–12833.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.*, *99*, 195–231.

Squire, L. R. (2004). The legacy of patient H. M. for neuroscience. *Neuron*, *61*, 6–9.

Squire, L. R., & Bayley, P. J. (2007). The neuroscience of remote memory. *Curr. Opin. Neurobiol.*, *17*, 185–196.

Squire, L. R., & Knowlton, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 765–779). Cambridge, MA: MIT Press.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.*, *27*, 279–306.

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nat. Rev. Neurosci.*, *8*, 872–883.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*, 1380–1386.

Squire, L. R., Zola-Morgan, S., & Chen, K. S. (1988). Human amnesia and animal models of amnesia: Performance of amnesic patients on tests designed for the monkey. *Behav. Neurosci.*, *102*, 210–221.

Stark, C. E., & Squire, L. R. (2000). Intact visual perceptual discrimination in humans in the absence of perirhinal cortex. *Learn. Memory*, *7*, 273–278.

Steinvorth, S., Levine, B., & Corkin, S. (2005). Medial temporal lobe structures are needed to re-experience remote autobiographical memories: Evidence from H.M. and W.R. *Neuropsychologia*, *43*, 479–496.

Teng, E., Stefanacci, L., Squire, L. R., & Zola, S. M. (2000). Contrasting effects on discrimination learning after hippocampal lesions and conjoint hippocampal-caudate lesions in monkeys. *J. Neurosci.*, *20*, 3853–3863.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301–306.

Wais, P. E., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron*, *49*, 459–466.

Warrington, E. K. (1996). Studies of retrograde memory: A long-term view. *Proc. Natl. Acad. Sci. USA*, *93*, 13523–13526.

Warrington, E. K., & Taylor, A. M. (1973). Immediate memory for faces: Long- or short-term memory? *Q. J. Exp. Psychol.*, *25*, 316–322.

Whitlock, J. R., Sutherland, R. J., Witter, M. P., Moser, M. B., & Moser, E. I. (2008). Navigating from hippocampus to parietal cortex. *Proc. Natl. Acad. Sci. USA*, *105*, 14755–14762.

Wixted, J. T., & Squire, L. R. (2004). Recall and recognition are equally impaired in patients with selective hippocampal damage. *Cogn. Affective Behav. Neurosci.*, *4*, 58–66.

Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, *12*, 323–339.

Yonelinas, A. P., Kroll, N. E., Quamme, J. R., Lazzara, M. M., Sauve, M. J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nat. Neurosci.*, *5*, 1236–1241.

47 Reconsolidation: A Possible Bridge between Cognitive and Neuroscientific Views of Memory

KARIM NADER

ABSTRACT The field of reconsolidation is one of the fastest growing fields in memory research. Students of memory find themselves in an extremely exciting period because memory research is beginning to be revealed at the neurobiological level as a fundamentally dynamic process. A neurobiological model of memory is emerging that can accommodate the dynamic nature of memory revealed in a long tradition of cognitive-oriented studies of human memory (Bartlett, 1932). This chapter will briefly address the history of consolidation and reconsolidation. It will describe the basis for which a reconsolidation phenomenon is thought to exist and address some central issues and unresolved problems of memory reconsolidation.

When students are asked to give analogies of how the brain processes memories, they often suggest that memories are like pictures stored in a filing cabinet. Remembering is suggested to consist of pulling the correct files from the brain for examination and then filing them again, unchanged, back into the storage banks of the brain. Memory—that is, the recall of previously stored memory contents—is thus simply faithful readout, just like opening a data file on a computer. Depending on the tradition in which the student was trained, this analogy is half wrong, but different parts of the analogy might be wrong in different traditions.

The cognitive tradition views memory as a reconstructive process, in which memories and their content are subject to change. This view was first developed by Bartlett in his seminal book *Remembering* (1932). He suggested that each (episodic) memory recall represents essentially a re-creation, based on one's current assumptions and beliefs about the world (schemata). A large body of work now supports this initial claim, indicating that under several conditions, memories can be easily corrupted (Loftus, 1997; Schacter, 1999). False memory paradigms can instill false memories within minutes, and subjects who are sensitive to these effects are often amazed and sometimes alarmed by how easily their memories can be manipulated without their consciously noticing the modifications. These and similar demonstrations established that remembering is not akin to a passive readout of a stored file; rather it is a reconstructive process in which new information is combined with current and previous experiences.

The physiological tradition, along with what is now referred to as systems neuroscience, somewhat limits the scope of the effects of malleability documented in cognitive psychology. Memory models developed in physiological psychology suggest that memories may only be substantially manipulated during a transient period of instability that follows their initial acquisition. It is assumed that memories initially exist in an unstable (labile) state and then become stable (consolidated) over time. This assumption explains why, if you are trying to remember a phone number and you are distracted within a few minutes of learning the phone number, you will likely forget the number in part or totally because the memory for the number was still in an unstable state when the interference occurred. If the same distraction were to happen 24 hours after the initial learning, chances are that the memory for the phone number would not be affected by the same distracter. Again, distracters are only effective shortly after new learning as only then is memory in an unstable state, whereas, after a few hours have passed, memory would have been stored, or "wired," into the brain, resistant to distracters. In other words, once memory is stabilized, it becomes fixed and cannot be changed as easily as it can shortly after learning (McGaugh, 2000). The processes involved in memory "fixation" have been the main focus in a highly successful research program in the physiological tradition. How do memories become fixed over time in the brain? What are the cellular and molecular mechanisms mediating this transformation from a labile to a consolidated state?

It is obvious that there are only small areas of overlap between these two traditions. On the one hand, cognitive psychology reminds us that our memories are not snapshots of the past but elaborate reconstructions. On the other hand, many physiologists and neuroscientists study the mechanisms

KARIM NADER Psychology Department, McGill University, Montreal, Quebec

of memory as a unidirectional process of fixation. Both views are based on a broad basis of empirical evidence—so how can they be reconciled? The recent rediscovery of reconsolidation—the phenomenon that seemingly consolidated, that is, stable, memories can return for a short time to a labile state after they have been reactivated—may begin to create points of overlap between the two traditions.

Consolidation theory

One of the defining features of memory is that it is a time-dependent process. New memories initially exist in an unstable state (called a labile state) and then over time become more stable, entering a state in which they are resistant to disruptions (Ebbinghaus, 1885; Müller & Pilzecker, 1900; Ribot, 1881). The transformation of memory from an unstable to a stable state is what theories of consolidation try to explain. As in any field in which scientists strive to create reductive models of a phenomenon, terminology is often applied in various levels of analysis, leading to some confusion when used in a different context. "Consolidation" is no exception. The term is used in a variety of ways that differ depending on levels of analysis (e.g., McGaugh, 2000; Spear & Mueller, 1984; Squire, 1992). Thus it would be helpful to explicitly differentiate between two ways in which the term is commonly used today. *Systems* consolidation posits that hippocampus-dependent memory becomes, over years in humans or weeks in rodents, hippocampus-independent (Scoville & Milner, 1957; Squire & Alvarez, 1995). *Cellular* (or synaptic) consolidation refers to a universal property of neurons, namely, the time-dependent stabilization of changes in synaptic efficacy, which represent the neural substrate of memory, within hours following acquisition (Dudai, 2004; Kandel, 2001).

Empirical evidence for a cellular consolidation process is derived from three lines of evidence. First, performance can be impaired if brain or neuronal function is compromised with manipulations such as electroconvulsive shock or protein synthesis inhibitors (Duncan, 1949; Flexner, Flexner, & Stellar, 1965). Second, performance can be impaired by new competing learning (Müller & Pilzecker, 1900). Third, performance can be enhanced by a variety of drugs that in general terms "stimulate" brain function (McGaugh & Krivanek, 1970). In order for any of these manipulations to be effective they must be administered shortly after new learning but not after a delay of minutes to hours after learning. Data from these lines of evidence forms the foundation of contemporary consolidation theory (Glickman, 1961; McGaugh, 1966). It became increasingly clear that memory is not stored immediately in the brain but that this process takes time and changes some aspects of the participating neurons. The initial labile state became synonymous with short-term memory (STM), and when the memory was resistant to challenge, it was referred to as being consolidated, or "fixed," and thus became long-term memory (LTM). It is important to note that "labile" is conceptually differentiated from the strength of the memory; that is, it is not implied that a labile memory is somehow weaker than a stable one. For example, behavioral responding when the memory is labile (STM) is often comparable to when the memory is in a consolidated state (LTM) (see figure 47.1). One implication of consolidation theory is that consolidation is unidirectional. Once a memory is "fixed" in the brain, it will remain as such (Glickman, 1961; McGaugh, 1966).

In typical experiments studying consolidation, animals are initially trained on a new task, and then a treatment is administered that affects a molecule or molecular pathway deemed critical for consolidation. Powerful predictions can be made about when a behavioral effect should be seen by exploiting the time-dependent nature of memory-consolidation processes. If the treatment specifically impairs consolidation, then it is expected that that the behavioral impairment should reflect the time-dependent nature of consolidation: an impairment (or improvement, depending on the kind of treatment administered) should only be seen when the intervention occurred within the few hours memory was in the labile state, that is, STM, but not after it had entered LTM. Thus the operational definition of consolidation impairment is spared STM but impaired LTM (Dudai, 2004; McGaugh, 2000). It is difficult to imagine an alternative model that would predict a time-dependent impairment with these temporal characteristics. Alternative explanations of the time-dependent vulnerability of memory to interference, such as changes in memory strength, in retrievability, or in the content of the memory, predict that treatment effects should have immediate repercussions; that is, STM should be affected already, which is not the case. For these reasons, the consolidation model almost reached the status of a fact.

Studies using modern neurobiological tools at the genetic, molecular, cellular, and systems levels have led to the same basic result pattern—intact STM and impaired LTM. These approaches demonstrated across species and learning systems that, in order for a memory trace to become consolidated, new proteins and RNA must be produced (Kandel, 2001). Departing from the axiom that the unit of memory is a change in the strength of a synapse (Hebb, 1949), the underlying assumption of these investigations is that new proteins contribute to synapse strengthening by either increases in postsynaptic excitatory receptors (Malinow & Malenka, 2002), new synapse formation (Bailey & Kandel, 1993), or the increase in release of presynaptic neurotransmitters (T. V. Bliss & Collingridge, 1993) that stabilize the memory trace over longer time periods. Two of the main mechanisms of these changes are long-term potentiation and long-term depression (T. V. P. Bliss & Lomo, 1973; Martin, Grimwood, & Morris, 2000). These cellular changes, like

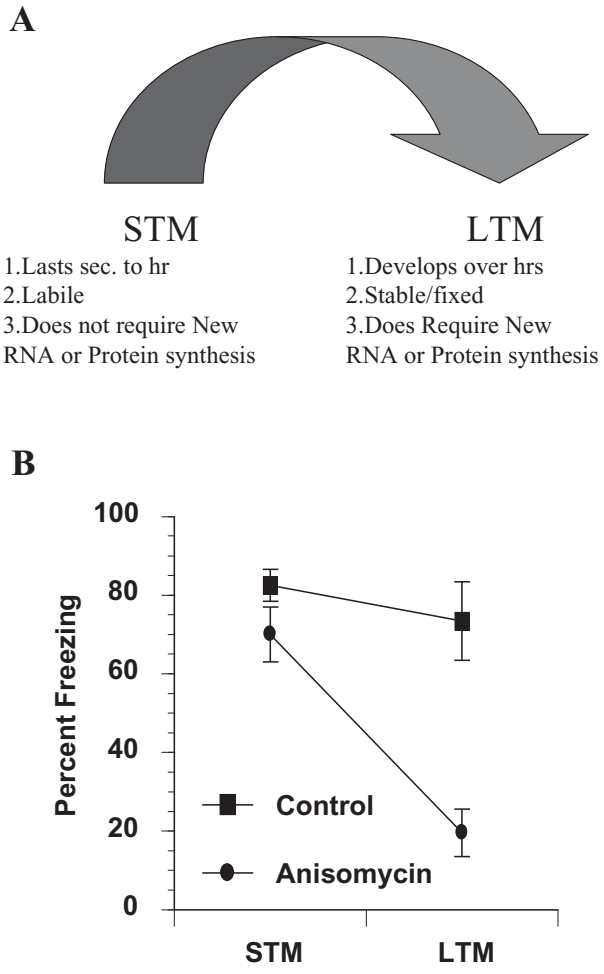


FIGURE 47.1 (*A*) A schematic of synaptic consolidation theory. Short-term memories (STM) are considered to be unstable or labile. They are thought to last on the order of minutes to hours. The currently dominant view is that STM expression does not require any new RNA or protein synthesis. Long-term memories are thought to develop over a few hours. They require new protein and RNA synthesis in order to become stable. (*B*) Data from Schafe, Nader, Blair, and LeDoux (2001) demonstrating that posttraining infusions of the protein synthesis inhibitor anisomycin into the basolateral amygdala blocks consolidation of auditory fear conditioning. Note that the data meet the operational definition of a consolidation blockade, intact STM and impaired LTM.

memory, have different phases: an early phase that is sensitive to inhibition of protein synthesis and a later phase that is not (Goelet, Castellucci, Schacher, & Kandel, 1986).

Reconsolidation

In 1968, Misanin, Miller, and Lewis conducted a clever experiment to test the assumption of unidirectionality, one of the main tenets of consolidation theory, which suggests that once memory is consolidated, it remains fixed in the brain. The authors first demonstrated that 24 hours after training the memory had become insensitive to electroconvulsive shock (ECS), which impaired memory when applied shortly after training, suggesting that the memory was consolidated about one day after learning. However, if they presented the animals 24 hours after training with a reminder just prior to ECS, amnesia would follow (Misanin et al., 1968). According to consolidation theory, as the memory had already become insensitive to ECS, and thus consolidated, it should also remain so following memory retrieval. The authors suggested that transformation of a memory from an unstable to a stable state occurred not only for new memories but also for consolidated memories after they had been recalled. This effect, which was originally called "cue-induced amnesia," led to a large number of studies. As with many (new) phenomena, the effect was not replicated in some paradigms (Dawson & McGaugh, 1969; Gold & King, 1972; Squire, Slater, & Chace, 1976). Nevertheless, it was found across a range of species and amnesic treatments (Lewis, 1979; Spear & Mueller, 1984), suggesting that the negative findings may point to certain conditions under which the phenomenon does not occur.

The initial studies suggested that reactivation of a memory induced a state of lability such that memory again had to be stabilized over time. Using the standards of consolidation theory and applying them to cue-induced amnesia, reactivation induces a time window during which performance can be affected by one of three different manipulations. First, performance can be impaired if brain or neuronal function is compromised (Misanin et al., 1968). Second, performance can be impaired by new competing learning (Gordon, 1977a). Third, performance can be enhanced by stimulants (Gordon, 1977b). When administered shortly after the reminder but not after a delay of minutes or hours, these manipulations affect performance. In addition, the manipulations are not effective when the memory is not reactivated (Misanin et al., 1968). Thus reactivation of a consolidated memory induces another state of transient lability that affords a process of time-dependent memory stabilization.

For a variety of reasons, however, cue-induced amnesia never became part of the dominant physiological tradition of memory. Some of the original scientists who demonstrated the effect (Land, Bunsey, & Riccio, 2000; Przybyslawski & Sara, 1997) and others (Rodriguez, Rodriguez, Phillips, & Martinez, 1993) continued to periodically publish reconsolidation results, but these results were met with indifference by the field, which had settled on consolidation theory as the defining paradigm of memory research.

Reviving Reconsolidation

Initially unaware of the early literature on "cue-induced amnesia," we completed a series of experiments that were aimed at testing whether reactivation of consolidated

memories would cause them to return to a labile state from which they had to be restabilized. We found that the role of protein synthesis in consolidation is *recapitulated* during reconsolidation (Nader, Schafe, & LeDoux, 2000a). These experiments used a fear-conditioning paradigm, for which the circuitry mediating learning is relatively well established (LeDoux, 2000). The most commonly used protein synthesis inhibitor, anisomycin, was *infused directly* into the lateral and basal nucleus of the amygdala (LBA), which is a site believed to mediate the learning and consolidation of fear-conditioned memories (LeDoux, 2000; Maren, 2001; Schafe & LeDoux, 2000). Specifically, infusions of anisomycin immediately after training had no effect on an STM retention test 4 hours after training but impaired performance on an LTM test at 24 hours after training (Schafe & LeDoux, 2000) (figure 47.1*B*). This finding suggests that the auditory fear memory is consolidated in the LBA.

Using this same logic to test for reconsolidation, rats were infused with anisomycin into the LBA immediately after reactivation and tested 4 hours later (postreactivation STM, PR-STM) and again at 24 hours after reactivation (postreactivation LTM, PR-LTM). These were analogues to STM and LTM tests, respectively. We predicted that if reactivation of a consolidated memory initiated a second time-dependent memory process, then anisomycin infusions into the LBA should have no effect on PR-STM but should impair PR-LTM performance. This was the pattern of results that we found (figure 47.2*B*; Nader et al., 2000a). Furthermore, in the absence of memory reactivation, the same anisomycin injections did not induce amnesia, indicating that the memory was consolidated at that time. Based on these and other findings, we concluded that reactivation of a consolidated memory initiates a second time-dependent memory process that requires protein synthesis in the amygdala in order to be restabilized. The restabilization process is now called reconsolidation. The term was used at least as early as 1973 (Spear, 1973) and has been used in more recent times by Przybyslawski and Sara (1997) and Rodriguez and colleagues (1993). Again it is important to point out that making a memory "labile" is not consistent with it being weakened in any way, as suggested, for example, by Rudy, Biedenkapp, Moineau, and Bolding (2006), who argue that lability requires that memories be eliminated and then put back into the brain. This position is clearly at odds with the data and the proposed explanations of the phenomenon (Nader et al.).

Our findings have again led to a large number of studies, and again the evidence is much the same as after the original description of the phenomenon by Misanin and colleagues (1968). A time-dependent behavioral impairment has now been demonstrated across a variety of tasks (figure 47.3), such as object recognition (Kelly, Laroche, & Davis, 2003), incentive learning (Sangha, Scheibenstock, & Lukowiak, 2003), inhibitory avoidance (Anokhin, Tiunova, & Rose, 2002),

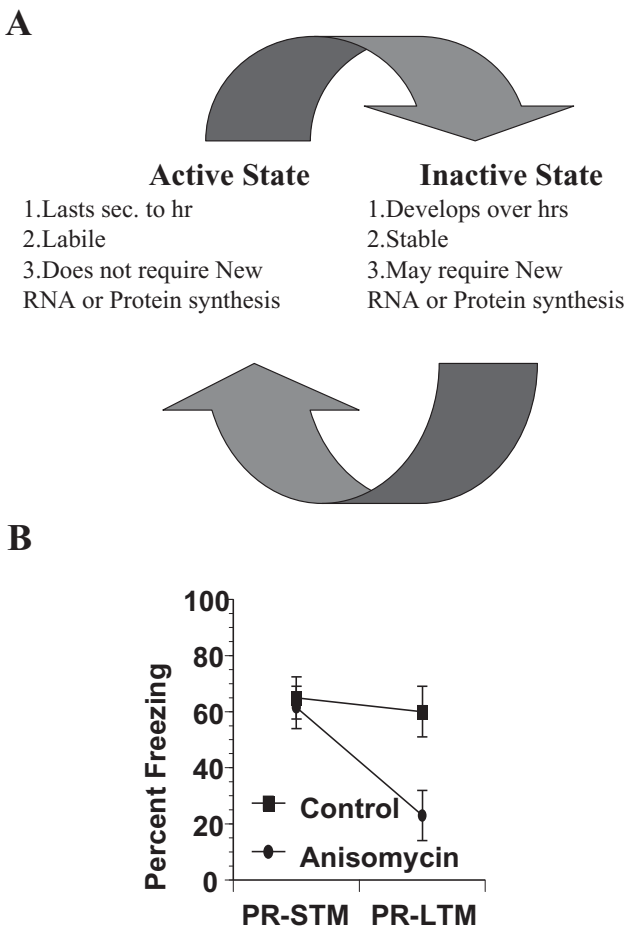


FIGURE 47.2 (*A*) A schematic of Donald Lewis's theory of memory processing. Active memories (AM) are considered to be unstable or labile. They are thought to last on the order of minutes to hours. Currently the evidence suggests that its expression does not require any new RNA or protein synthesis. Inactive memories (IM) are thought to develop over a few hours. Currently the data suggest that they require new protein and RNA synthesis in order to become stable. (*B*) Data from Nader, Schafe, and LeDoux (2000a) demonstrating that postreactivation infusions of the protein synthesis inhibitor anisomycin into the basolateral amygdala blocks reconsolidation of auditory fear conditioning. Note that the data meet the operational definition of a consolidation blockade, intact PR-STM and impaired PR-LTM.

motor sequence learning (Walker, Brakefield, Hobson, & Stickgold, 2003), appetitive conditioning (Wang, Ostlund, Nader, & Balleine, 2005), episodic memories (Forcato et al., 2007; Hupbach, Gomez, Hardt, & Nadel, 2007), and memories of drug reward (Lee, Di Ciano, Thomas, & Everitt, 2005; C. Miller & Marshall, 2005). A variety of different amnesic treatments have been shown to be effective in blocking reconsolidation, such as targeted protein (Nader et al., 2000a) or RNA synthesis inhibition (Duvarci, Nader, & LeDoux, 2008; Sangha et al., 2003), pharmacological inhibition of kinase activity (Kelly et al.), protein knockout mice (Bozon, Davis, & Laroche, 2003), inducible knockout mice (Kida et al., 2002),

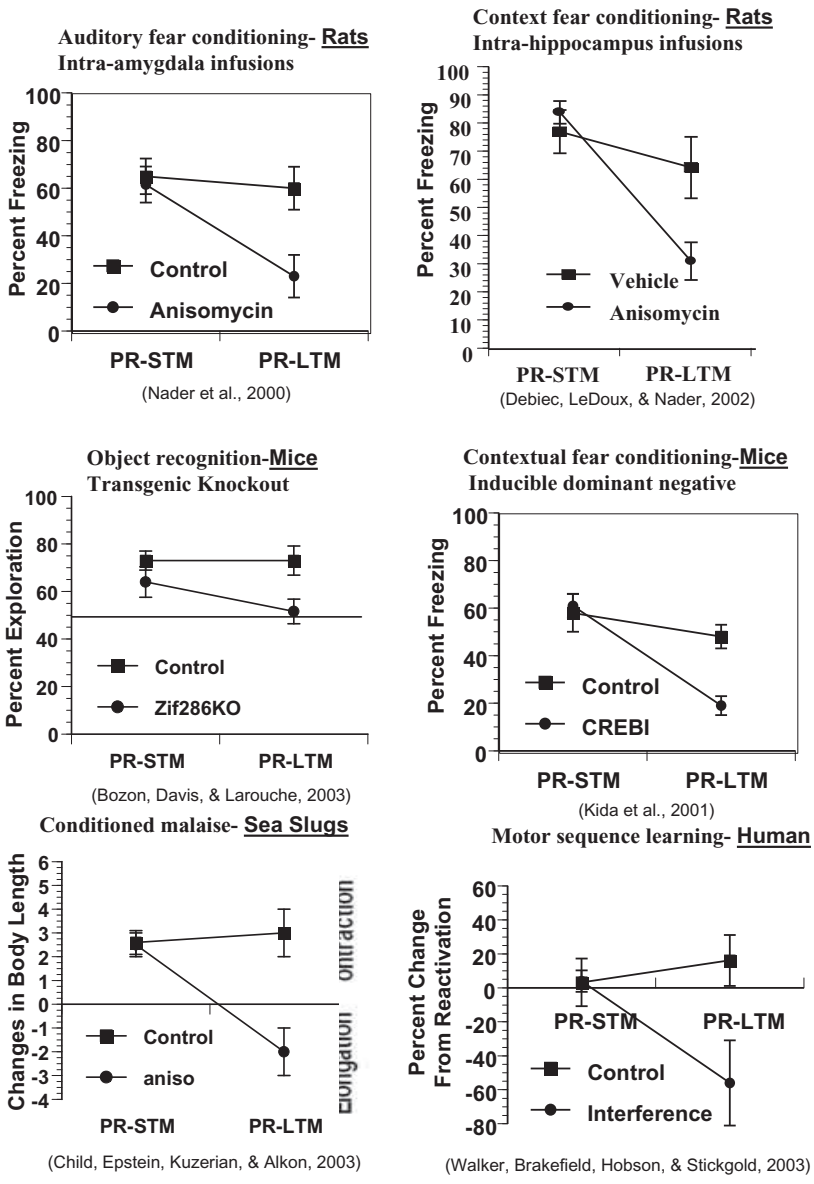


FIGURE 47.3 Representative data from some of the studies reporting a reconsolidation impairment. Notice that they all show intact postreactivation STM (PR-STM) and impaired postreactivation LTM (PR-LTM). Labels above each graph refer to the paradigm and the nature of the amnesia treatment. Below each are the citations for the data.

beta-adrenergic antagonists (Przybyslawski & Sara, 1997), and simply an interference by new learning (Forcato et al.; Hupbach et al.; Walker et al., 2003). In addition, reconsolidation has been reported across a broad spectrum of species, such as snails (Sangha et al., 2003), sea slugs (Child, Epstein, Kuzirian, & Alkon, 2003), crabs (Pedreira, Perez-Cuesta, & Maldonado, 2002), chicks (Anokhin et al.), mice (Kida et al.), rats (Nader et al., 2000a), and most importantly humans (Brunet et al., 2008; Forcato et al.; Hupbach et al.; Walker et al.). While most of the aforementioned studies have focused on blockade of reconsolidation, there is also evidence that reconsolidation can be enhanced by postreactivation increases in kinase activity (Tronson, Wiseman, Olausson, & Taylor, 2006).

CONSTRAINTS ON RECONSOLIDATION The data set shows that reconsolidation can be found across a large number of paradigms and species, suggesting that reconsolidation is a fundamental phenomenon. However, it is apparently not ubiquitous, as some have failed to demonstrate it in the late phase of instrumental conditioning or memory of context (Biedenkapp & Rudy, 2004; Hernandez, Sadeghian, & Kelley, 2002). This fact would indicate that perhaps not all memories undergo reconsolidation. Furthermore, there is a large amount of exciting work demonstrating that the ability of a memory to undergo reconsolidation can be influenced by certain experimental parameters. For example, some older memories may be more resistant to reconsolidation (Eisenberg & Dudai,

2004; Milekic & Alberini, 2002; Suzuki et al., 2004). In these studies, the authors report that the ability of postreactivation treatments to induce amnesia decreases with time between training and reactivation. This finding would suggest that old memories may not undergo reconsolidation.

Another suggested constraint on reconsolidation is that if the reminder is sufficiently robust to induce extinction of the memory, extinction prevents the memory from undergoing reconsolidation in many paradigms (Eisenberg, Kobilo, Berman, & Dudai, 2003; Pedreira & Maldonado, 2003; Sangha, Scheibenstock, Morrow, et al., 2003; Suzuki et al., 2004).

In collaboration with Joseph LeDoux's group, we wanted to test whether it mattered if a memory was directly or indirectly reactivated in order to undergo reconsolidation. Reductionists often talk about memories as if they exist in isolation; scientists test consolidation of "this" or "that" memory, training animals in a variety of tasks and treating the learning as if this were the only thing animals ever experience. Therefore, from a mnemonic perspective, the experimental animal's arrival in the lab is ground zero. From then onward, the animals are welcome to have memories. In the real world, of course, memories are richly associated with other memories (Tulving, 2002). We wanted to test whether the manner in which a memory was reactivated, directly or indirectly, influenced a memory's ability to undergo reconsolidation. We found that memories that were directly reactivated underwent reconsolidation; however, memories that were indirectly reactivated did not (figure 47.4) (Debiec, Doyere, Nader, & LeDoux, 2006). This constraint has yet to be tested in other paradigms, so it remains unclear whether it will generalize to other tasks and memory systems.

For all of the suggested constraints on reconsolidation there are published exceptions to the rule. For example, extinction of an auditory fear memory does not prevent the memory from undergoing reconsolidation in the amygdala (Duvarci, Mamou, & Nader, 2006). In honeybees, extinction of an appetitive response was necessary to *induce* the memory to undergo reconsolidation (Stollhoff, Menzel, & Eisenhardt, 2005). For age of memories, a large number of studies have shown that even very old memories can undergo reconsolidation (Brunet et al., 2008; Debiec, LeDoux, & Nader, 2002; Lee et al., 2005). Indeed, in one important study, patients with posttraumatic stress disorder had reconsolidation of their traumatic memories reactivated and targeted with a beta antagonist. We reported that the strength of the traumatic memory decreased even though the average time between trauma and treatment was approximately 11 years (Brunet et al.). It is still too early to know why age and extinction prevent only some memories from undergoing reconsolidation. However, the conditions that constrain when a memory undergoes reconsolidation are a very exciting and dynamic area of current research.

Alternative Interpretations Given the richness of the data on reconsolidation mentioned previously, including the original findings, alternative interpretations do not readily explain the data set. In addition, because reconsolidation has been defined using the standards of consolidation, nonspecific interpretations are also applicable to the field of consolidation, challenging the latter in the same way as they purportedly challenge the former. The two main alternative interpretations, suggesting that the reconsolidation impairment represents nonspecific effects (including lesion) or that it represents a retrieval impairment, will be discussed in this section. For an extensive review of the issues surrounding reconsolidation, including state-dependent learning and facilitated extinction, see Nader (2003), Nader and Hardt (2009), and Nader and Wang (2006).

Recently it has been suggested that reconsolidation impairments represent nonspecific effects or lesions due to apoptosis (Rudy et al., 2006). In general, this interpretation has difficulty explaining the entire set of reconsolidation results. One has to assume that in all studies that show evidence for reconsolidation, lesions or nonspecific effects are responsible, despite the use of different amnesic treatments, tasks, and species, including new learning and inducible dominant negative CREB mutant mice.

In particular, there are three lines of evidence that make this interpretation of our data unlikely. First, anisomycin-treated animals show intact PR-STM, suggesting that no functional lesion is present 4 hours after reactivation and amnesic agent treatment. However, it is still possible that the behavioral impairments were due to a lesion that developed between the 4-hour PR-STM and 24-hour PR-LTM test. Speaking against this possibility is that anisomycin infusions 6 hours after reactivation had no effect on PR-LTM (Nader et al., 2000a). If anisomycin had produced a functional lesion, then the delayed infusion should also have caused an impairment during the PR-LTM test; however, that was not seen (Nader, Schafe, & LeDoux, 2000b). Second, a similar behavioral impairment is found when amnesic agents are used that target different points in the molecular cascade thought to be important for memory stabilization, such as an inhibitor of the extracellular signal-regulated kinase-mitogen-activated protein kinase (ERK-MAPK) U0126 (Duvarci, Nader, & LeDoux, 2005) and RNA synthesis inhibitors (alpha-amanatin and DRB) (Duvarci et al., 2008).

The question of whether experimental amnesia, induced by blocking novel or reactivated memories, represents an impairment of the consolidation (storage) process or an impairment in retrieval of an otherwise sufficiently consolidated memory has never been resolved (Gold & King, 1974; R. Miller & Springer, 1974). There were a number of reasons for this stalemate. One was that both views of amnesia could explain any results produced within the dominant paradigm used to study the nature of amnesia, the recovery-

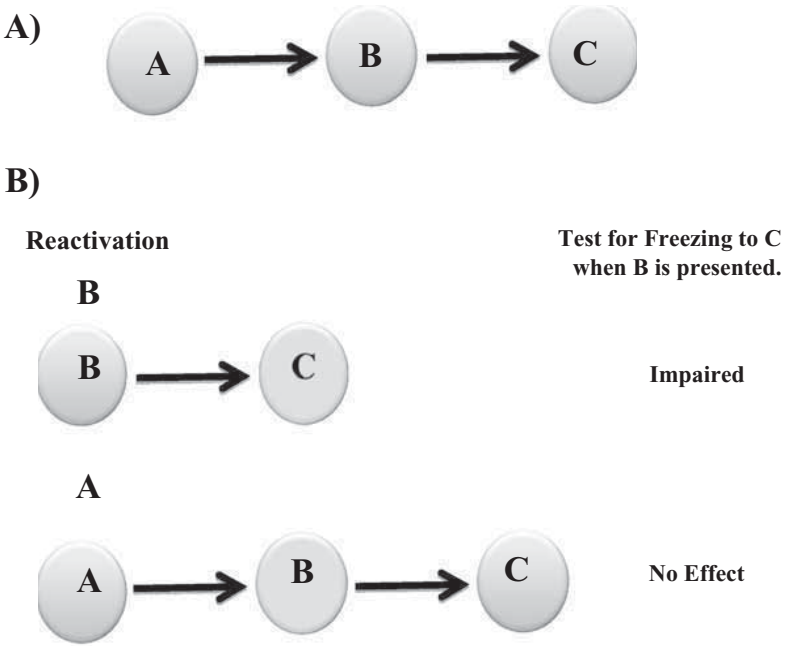


FIGURE 47.4 (*A*) Cartoon of the chain of associations created to test whether being reactivated in an indirect manner changed a memory's propensity to undergo reconsolidation (Doyere, Debiec, Monfils, Schafe, & LeDoux, 2007). (*B*) Presenting B during the reactivation *directly* elicits the B → C association that undergoes reconsolidation. Amnesic treatments would then induce an impairment in responding if B was presented on test. Indirect reactivation of the same association by presenting A during reactivation causes the directly reactivated A → B memory to undergo reconsolidation but not the indirectly reactivated association, B → C. Therefore, amnesic agents had no effect on responding to B.

from-amnesia paradigm. Furthermore, within this paradigm the consolidation-impairment view of amnesia only makes negative predictions, which do not prove that a memory is not present (Nader & Wang, 2006; Squire, 2006). For this reason, even at the present time, there are completely opposed views on the nature of experimental amnesia (for examples of the discordance on the issue see the special section in *Learning and Memory*, *13*[5], 2006).

Consequently, we cannot address whether amnesia induced by postreactivation treatments is due to impairing reconsolidation or retrieval processes. However, what we can do is compare this amnesia to amnesia for new memories and apply the accepted standards in the field of consolidation to conclude that the former deficit is due to impaired memory storage. At the behavioral level, amnesia for new memories (Anokhin et al., 2002; Lattal & Abel, 2004; Quartermain & McEwen, 1970; Squire & Barondes, 1972) and reactivated memories (Anokhin et al.; Lattal & Abel; Quartermain & McEwen, 1970; Squire & Barondes, 1972) are similar, as they can both show recovery. Recovery from amnesia, however, can be consistent with both storage and retrieval-impairment views of amnesia (Nader & Wang, 2006; Squire, 2006).

Because the recovery-from-amnesia paradigm did not resolve the issue, the field has also examined whether the molecular and cellular changes that occur during the post-training stabilization, or consolidation, period are lost in amnesic animals (Squire, 2006). For example, in aplysia the consolidation of long-term facilitation is accompanied by an increase in synapse number (Bailey & Kandel, 1993). However, in amnesic preparations, in which a memory deficit was induced by inhibiting CREB phosphorylation, increase in synapse number is not observed (Bailey & Kandel, 1993). Such data are considered exclusively as evidence supporting the view that amnesia represents impairment of memory storage (Squire, 2006). While this is certainly a very positive step to advancing the issue, there are still theories of memory processing at the behavioral level (Lewis, 1979; R. Miller & Marlin, 1984; Spear, 1973) that may provide alternative interpretations of this kind of data that remain consistent with the behavioral impairment being a retrieval failure (Nader & Wang, 2006).

A few scientists have begun to examine what happens to established molecular and cellular correlates of LTM when reconsolidation is blocked. Importantly, long-term potentiation in its late phase (L-LTP) can undergo a reconsolidation-like process (Fonseca, Nagerl, & Bonhoeffer, 2006). In addition, learning-induced increases in field potentials in the amygdala are decreased when reconsolidation is blocked for that memory (Doyere, Debiec, Monfils, Schafe, & LeDoux, 2007). Molecular correlates of LTM have also been shown to return toward baseline levels when reconsolidation is blocked (C. Miller & Marshall, 2005; Rose & Rankin, 2006; Valjent et al., 2006), as if the brain area targeted by the amnesic agent reduced the plasticity in those circuits. Within

the physiological tradition, reversal of cellular and molecular signatures of LTM is the accepted standard for concluding that amnesia is a storage impairment (Squire, 2006). Therefore these studies must imply that reconsolidation deficits are due to impairments in memory restorage.

As briefly mentioned previously, a number of alternative interpretations have been proposed for the reconsolidation effect, including nonspecific effects such as lesions (Rudy et al., 2006), transient retrieval impairment (Cahill, McGaugh, & Weinberger, 2001; McGaugh, 2004; Squire, 2006), facilitated extinction (Fischer, Sananbenesi, Schrick, Spiess, & Radulovic, 2004; Myers & Davis, 2002), internal reinforcement (Eisenhardt & Menzel, 2007), and state-dependent learning (Millin, Moody, & Riccio, 2001). As reconsolidation has been found across such a wide range of species and amnesic effects, and both behavioral impairments and *enhancements* have been reported, these nonspecific interpretations have difficulty explaining all findings (Alberini, 2005; Nader, 2003). Indeed, none of these alternative interpretations predict time-dependent behavioral effects. They predict impairments in *both* PR-STM and PR-LTM, and none would predict enhancements in performance. Furthermore, we know that there are constraining conditions that can sometimes determine whether amnesia would be induced. For example, post-reactivation infusions of anisomycin impair the performance of recent but not remote memories (Suzuki et al., 2004), but nonspecific interpretations predict that under all conditions amnesia would be observed. Last, it is important to restate that the reconsolidation interpretation is derived from the standard definitions developed in the consolidation field. Therefore any alternative interpretation of the reconsolidation impairment will generalize to the evidence on which the conclusion for the existence of a consolidation process rests.

Demonstrations of the reconsolidation effect, as Lewis initially pointed out, cannot be explained by consolidation theories. According to consolidation theories, the lability of a memory ceases once it enters LTM. Reconsolidation studies clearly show that reactivation of a consolidated memory can induce another time-dependent stabilization process. Lewis proposed a new model of memory processing that can incorporate findings of both consolidation and reconsolidation. He suggested that a memory can exist in two states—an active or an inactive state (Lewis, 1979). Active memories comprise either new memories or previously consolidated and now reactivated memories. Over time, memories in the active state stabilize to an inactive state. A future reminder may then return the memory again to an active state.

Systems Reconsolidation When speaking about consolidation and the hippocampus, a time-dependent memory-transformation process—systems consolidation—is implied that is assumed to last weeks (rodents) and years (humans) (Dudai & Morris, 2000). First described by Scoville and Milner (1957), the hypothesis states that the hippocampus plays a transient role in memory processing, such that recent memories are hippocampus-dependent while remote memories are not (Anagnostaras, Gale, & Fanselow, 2001; Eichenbaum, Otto, & Cohen, 1994; McClelland, McNaughton, & O'Reilly, 1995; Squire & Alvarez, 1995). Whether the hippocampus stays involved or stops being involved in remote memories has become the subject of a lively debate in recent years, and, as an alternative to the standard view articulated in systems consolidation, multiple-trace theory was proposed (Nadel, Samsonovich, Ryan, & Moscovitch, 2000). However, given the fact that systems consolidation still remains the dominant view, especially in the physiological tradition of memory research, the results concerning reconsolidation will be discussed only within this framework.

The first suggestion that something akin to systems reconsolidation (a remote memory that returns to being hippocampus-dependent with memory reactivation) could occur was a study by Riccio's group (Land, Bunsey, & Riccio, 2000). Using an avoidance task, they demonstrated that lesions of the hippocampus affected behavior 1 but not 30 days after training. However, lesions administered after 30 days became effective if the memory was reactivated before. This finding demonstrated that memory reactivation could cause the memory to become hippocampus-dependent again. However, because their task required the hippocampus for retrieval and not acquisition, it did not directly speak to the issue of systems reconsolidation.

We have demonstrated that hippocampus-dependent memories undergo both cellular and systems reconsolidation (Debiec et al., 2002) (figure 47.5). Our study used a contextual fear-conditioning paradigm in conjunction with either targeted infusions of the protein-synthesis inhibitor anisomycin into the dorsal hippocampus or lesions of this structure. Consistent with previous findings, lesions of the hippocampus 45 days after conditioning had no effect on the subsequent expression of contextual fear conditioning (Anagnostaras, Maren, & Fanselow, 1999; Kim & Fanselow, 1992). However, if the memory was reactivated for as little as 90 seconds prior to the same lesion, amnesia was observed and did not recover either with multiple testing protocols or spontaneously with time. Therefore memory reactivation of a remote memory can return it to a hippocampus-dependent state. Interestingly, this reactivated remote trace remained hippocampus-dependent for only two days. Thus, while the duration of systems consolidation (first retrograde amnesic gradient) is on the order of weeks (Anagnostaras et al., 1999; Kim & Fanselow, 1992) (up to 45 days in the present study), systems reconsolidation (second retrograde amnesic gradient) is on the order of days (here, two days). Further, a third gradient of comparable duration to the second was reported (Debiec et al., 2002).

The idea of systems reconsolidation has recently been supported by physiological findings demonstrating that

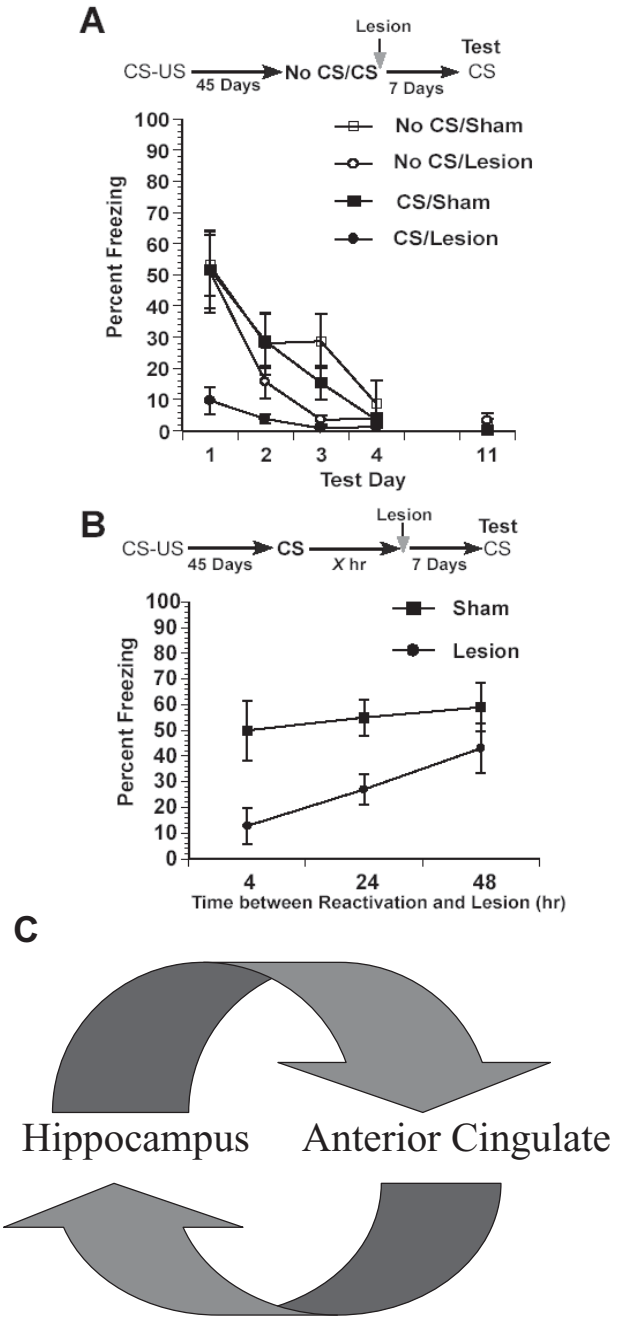


FIGURE 47.5 Systems reconsolidation in the hippocampus. Data from a contextual fear-conditioning paradigm demonstrating systems reconsolidation. Training (CS-US) consisted of 8 shock presentations in a conditioning chamber. (*A*) 45 days after conditioning, electrolytic lesions of the dorsal hippocampus (lesion) immediately after memory reactivation (CS) produced a significant impairment. Conversely, the same lesions had no effect when the reactivation session (no CS) was omitted, demonstrating that 45 days after conditioning the memory is independent of the hippocampus. Thus reactivation of a hippocampus-independent memory returns it to being hippocampus-dependent, an example of systems reconsolidation. (*B*) Reactivation of the remote memory returned it to being dependent on the hippocampus for less than 2 days. (*C*) Memory model of the hippocampus demonstrating both-systems reconsolidation. Over time the neocortex (possibly the anterior cingulate) becomes competent to mediate a simple response and might no longer need the hippocampus, at which point it is a remote memory (top arrow). Reactivation of the remote memory causes the cortical trace, which remains in the cortex, to require hippocampus feedback (bottom arrow) over the next 2 days. (From Debiec, LeDoux, & Nader, 2002.)

theta-phase synchronization between the lateral amygdala and the CA1 area of the hippocampus, which is lost for remote memories, returns the day after the reactivation of a remote memory (Narayanan, Seidenbecher, Sangha, Stork, & Pape, 2007). The authors suggest that this finding is evidence that reactivation of a remote memory returns it to a hippocampus-dependent state in which it again critically interacts with the lateral amygdala. In addition, some behavioral findings consistent with systems reconsolidation have been reported. It has been shown that contextual fear memories increase in their generalization to other contexts with age (Winocur, Moscovitch, & Sekeres, 2007). For young memories, animals were able to discriminate between a context associated with a foot-shock and a second distinct nonshock context by freezing more in the shock context. However, for remote memories, animals froze equally between the shocked and nonshocked contexts. The authors suggest that this result is more consistent with a transformation view of memory processing by the hippocampus, as opposed to a consolidation view. One implication of this finding is that when contextual memories are hippocampus-dependent, they permit discrimination of shocked and nonshocked contexts. Over time, when these memories putatively become independent of the hippocampus, animals lose the ability to discriminate between the two contexts. One study has shown that when remote memories are reactivated for 1 minute, then differential context-specific freezing is reinstated (Wiltgen & Silva, 2007). The authors suggest that this differential freezing may be due to systems reconsolidation that occurred when the memory returned to a hippocampus-dependent state. Converging evidence is being accumulated for systems consolidation.

COGNITIVE IMPLICATIONS OF RECONSOLIDATION The traditional physiological view of memory processing viewed memory as becoming wired into the brain over time and then remaining fixed (McGaugh, 1966). This "wiring" is typically discussed as being mediated by increases in postsynaptic receptors (Malinow & Malenka, 2002), growth of new synapses (Bailey & Chen, 1983), or the increase in release of neurotransmitters (T. V. Bliss & Collingridge, 1993). This physiological view is at odds with the cognitive tradition of memory, which views memory as being a *reconstructive* dynamic process. How then can memory be dynamic when it is wired into the brain? Part of the appeal of reconsolidation is that it provides a plausible

neurobiological mechanism for explaining some of the dynamic properties of memory. Indeed, based on the early studies on reconsolidation, Elizabeth Loftus suggested that cue-induced amnesia might be a mechanism underlying some false memories (Loftus & Yuille, 1984).

Perhaps the most straightforward predictions one can make for cognitive neuroscience is that if a specific impairment of reconsolidation is observed in humans, then intact PR-STM and impaired PR-LTM retention should be observed. In addition, any brain area activation that is associated with the presence of the memory should show return toward baseline only for the PR-LTM test. In contrast, the normal PR-STM test should be associated with activation. The amnesic agents could be either a pharmacological agent (Brunet et al., 2008), interference by new learning (Forcato et al., 2007; Hupbach et al., 2007; Walker et al., 2003), suppression (Anderson & Green, 2001), or transcranial magnetic stimulation (Duque et al., 2008). Because memory is one of few psychological time-dependent processes on this time scale, the intact PR-STM and impaired PR-LTM are very powerful tools to characterize both deficits of retention and their underlying substrates.

As mentioned earlier, some memories that undergo reconsolidation will do so only under some conditions. For example, age and extinction of some memories can make them more resistant to undergoing reconsolidation. Another interesting venue of research with regard to reconsolidation as a possible mediator of false memory would be to apply the parameters that control the occurrence of reconsolidation to false-memory studies. If reconsolidation must be present in order for false memory to occur, then false memory should only occur under the experimental conditions favorable to the occurrence of reconsolidation (for example, when the memory is young or directly reactivated).

Summary

Reconsolidation is beginning to provide areas of overlap between cognitive and physiological traditions of memory. There can be no question at this point that memories are fundamentally dynamic, as first explicitly demonstrated by Bartlett (1932). They are not snapshots of events that are passively read out; they are *constructive* in nature and always changing (Loftus & Yuille, 1984; Schacter, 1999; Tulving & Thomson, 1973). But there is also no doubt that memories stabilize over time (Glickman, 1961; McGaugh, 1966; Müller & Pilzecker, 1900). Models now exist across levels of analysis to describe the processes engaged during memory storage (Kandel, 2001; McGaugh, 2000). Reconsolidation, which has been shown to occur in species ranging from slugs to humans, is a mechanism that naturally endows neural systems with the property to both make memories more stable and then return them to unstable, malleable states in which they may be either strengthened, weakened, or changed—exactly the kind of properties one would want in a system to allow memories to change with future use.

REFERENCES

Alberini, C. M. (2005). Mechanisms of memory stabilization: Are consolidation and reconsolidation similar or distinct processes? *Trends Neurosci.*, *28*(1), 51–56.

Anagnostaras, S. G., Gale, G. D., & Fanselow, M. S. (2001). Hippocampus and contextual fear conditioning: Recent controversies and advances. *Hippocampus*, *11*(1), 8–17.

Anagnostaras, S. G., Maren, S., & Fanselow, M. S. (1999). Temporally graded retrograde amnesia of contextual fear after hippocampal damage in rats: Within-subjects examination. *J. Neurosci.*, *19*(3), 1106–1114.

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366–369.

Anokhin, K. V., Tiunova, A. A., & Rose, S. P. (2002). Reminder effects—reconsolidation or retrieval deficit? Pharmacological dissection with protein synthesis inhibitors following reminder for a passive-avoidance task in young chicks. *Eur. J. Neurosci.*, *15*(11), 1759–1765.

Bailey, C. H., & Chen, M. (1983). Morphological basis of long-term habituation and sensitization in Aplysia. *Science, 220*(4592), 91–93.

Bailey, C. H., & Kandel, E. R. (1993). Structural changes accompanying memory storage. *Annu. Rev. Physiol.*, *55*, 397–426.

Bartlett, F. C. (1932). *Remembering.* Cambridge, UK: Cambridge University Press.

Biedenkapp, J. C., & Rudy, J. W. (2004). Context memories and reactivation: Constraints on the reconsolidation hypothesis. *Behav. Neurosci.*, *118*(5), 956–964.

Bliss, T. V., & Collingridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature*, *361*(6407), 31–39.

Bliss, T. V. P., & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.*, *232*, 331–356.

Bozon, B., Davis, S., & Laroche, S. (2003). A requirement for the immediate early gene zif268 in reconsolidation of recognition memory after retrieval. *Neuron*, *40*(4), 695–701.

Brunet, A., Orr, S. P., Tremblay, J., Robertson, K., Nader, K., & Pitman, R. K. (2008). Effect of post-retrieval propranolol on psychophysiologic responding during subsequent script-driven traumatic imagery in post-traumatic stress disorder. *J. Psychiatr. Res.*, *42*(6), 503–506.

Cahill, L., McGaugh, J. L., & Weinberger, N. M. (2001). The neurobiology of learning and memory: Some reminders to remember. *Trends Neurosci.*, *24*(10), 578–581.

Child, F. M., Epstein, H. T., Kuzirian, A. M., & Alkon, D. L. (2003). Memory reconsolidation in hermissenda. *Biol. Bull.*, *205*(2), 218–219.

Dawson, R. G., & McGaugh, J. L. (1969). Electroconvulsive shock effects on a reactivated memory trace: Further examination. *Science*, *166*(3904), 525–527.

Debiec, J., Doyere, V., Nader, K., & LeDoux, J. E. (2006). Directly reactivated, but not indirectly reactivated, memories undergo reconsolidation in the amygdala. *Proc. Natl. Acad. Sci. USA*, *103*(9), 3428–3433.

Debiec, J., LeDoux, J. E., & Nader, K. (2002). Cellular and systems reconsolidation in the hippocampus. *Neuron*, *36*(3), 527–538.

Doyere, V., Debiec, J., Monfils, M. H., Schafe, G. E., & LeDoux, J. E. (2007). Synapse-specific reconsolidation of distinct fear memories in the lateral amygdala. *Nat. Neurosci.*, *10*(4), 414–416.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.*, *55*, 51–86.

Dudai, Y., & Morris, R. (2000). To consolidate or not to consolidate: What are the questions? In J. Bolhius (Ed.), *Brain, perception, memory: Advances in cognitive sciences* (pp. 149–162). Oxford, UK: Oxford University Press.

Duncan, C. P. (1949). The retroactive effect of electroconvulsive shock. *J. Comp. Physiol. Psychol.*, *42*, 32–44.

Duque, J., Mazzocchio, R., Stefan, K., Hummel, F., Olivier, E., & Cohen, L. G. (2008). Memory formation in the motor cortex ipsilateral to a training hand. *Cereb. Cortex*, *18*(6), 1395–1406.

Duvarci, S., Mamou, C. B., & Nader, K. (2006). Extinction is not a sufficient condition to prevent fear memories from undergoing reconsolidation in the basolateral amygdala. *Eur. J. Neurosci.*, *24*(1), 249–260.

Duvarci, S., Nader, K., & LeDoux, J. E. (2005). Activation of extracellular signal-regulated kinase-mitogen-activated protein kinase cascade in the amygdala is required for memory reconsolidation of auditory fear conditioning. *Eur. J. Neurosci.*, *21*(1), 283–289.

Duvarci, S., Nader, K., & LeDoux, J. E. (2008). A comparision of the sensitivities of consolidation and reconsolidation to RNA synthesis inhibition. *Learn. Memory, 15*(10), 747–755.

Ebbinghaus, M. (1885). *Über das Gedächtnis*. Leipzig: K. Buehler.

Eichenbaum, H., Otto, T., & Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behav. Brain Sci.*, *17*, 449–518.

Eisenberg, M., & Dudai, Y. (2004). Reconsolidation of fresh, remote, and extinguished fear memory in Medaka: Old fears don't die. *Eur. J. Neurosci.*, *20*(12), 3397–3403.

Eisenberg, M., Kobilo, T., Berman, D. E., & Dudai, Y. (2003). Stability of retrieved memory: Inverse correlation with trace dominance. *Science*, *301*(5636), 1102–1104.

Eisenhardt, D., & Menzel, R. (2007). Extinction learning, reconsolidation and the internal reinforcement hypothesis. *Neurobiol. Learn. Mem.*, *87*(2), 167–173.

Fischer, A., Sananbenesi, F., Schrick, C., Spiess, J., & Radulovic, J. (2004). Distinct roles of hippocampal de novo protein synthesis and actin rearrangement in extinction of contextual fear. *J. Neurosci.*, *24*(8), 1962–1966.

Flexner, L. B., Flexner, J. B., & Stellar, E. (1965). Memory and cerebral protein synthesis in mice as affected by graded amounts of puromycin. *Exp. Neurol.*, *13*(3), 264–272.

Fonseca, R., Nagerl, U. V., & Bonhoeffer, T. (2006). Neuronal activity determines the protein synthesis dependence of long-term potentiation. *Nat. Neurosci.*, *9*(4), 478–480.

Forcato, C., Burgos, V. L., Argibay, P. F., Molina, V. A., Pedreira, M. E., & Maldonado, H. (2007). Reconsolidation of declarative memory in humans. *Learn. Memory*, *14*(4), 295–303.

Glickman, S. (1961). Perseverative neural processes and consolidation of the memory trace. *Psychol. Bull.*, *58*, 218–233.

Goelet, P., Castellucci, V. F., Schacher, S., & Kandel, E. R. (1986). The long and short of long-term memory—A molecular framework. *Nature*, *322*(July), 419–422.

Gold, P. E., & King, R. A. (1972). Amnesia: Tests of the effect of delayed footshock-electroconvulsive shock pairings. *Physiol. Behav.*, *8*(5), 797–800.

Gold, P. E., & King, R. A. (1974). Retrograde amnesia: Storage failure versus retrieval failure. *Psychol. Rev.*, *81*(5), 465–469.

Gordon, W. C. (1977a). Similarities of recently acquired and reactivated memories in interference. *Am. J. Psychol.*, *90*(2), 231–242.

Gordon, W. C. (1977b). Susceptibility of a reactivated memory to the effects of strychnine: A time-dependent phenomenon. *Physiol. Behav.*, *18*(1), 95–99.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hernandez, P. J., Sadeghian, K., & Kelley, A. E. (2002). Early consolidation of instrumental learning requires protein synthesis in the nucleus accumbens. *Nat. Neurosci.*, *5*(12), 1327–1331.

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learn. Memory*, *14*(1–2), 47–53.

Kandel, E. R. (2001). The molecular biology of memory storage: A dialogue between genes and synapses. *Science*, *294*(5544), 1030–1038.

Kelly, A., Laroche, S., & Davis, S. (2003). Activation of mitogen-activated protein kinase/extracellular signal-regulated kinase in hippocampal circuitry is required for consolidation and reconsolidation of recognition memory. *J. Neurosci.*, *23*(12), 5354–5360.

Kida, S., Josselyn, S. A., de Ortiz, S. P., Kogan, J. H., Chevere, I., Masushige, S., et al. (2002). CREB required for the stability of new and reactivated fear memories. *Nat. Neurosci.*, *5*(4), 348–355.

Kim, J. J., & Fanselow, M. S. (1992). Modality-specific retrograde amnesia of fear. *Science*, *256*, 675–677.

Land, C., Bunsey, M., & Riccio, D. C. (2000). Anomalous properties of hippocampal lesion-induced retrograde amnesia. *Psychobiology*, *28*, 476–485.

Lattal, K. M., & Abel, T. (2004). Behavioral impairments caused by injections of the protein synthesis inhibitor anisomycin after contextual retrieval reverse with time. *Proc. Natl. Acad. Sci. USA*, *101*, 4667–4672.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.*, *23*, 155–184.

Lee, J. L., Di Ciano, P., Thomas, K. L., & Everitt, B. J. (2005). Disrupting reconsolidation of drug memories reduces cocaine-seeking behavior. *Neuron*, *47*(6), 795–801.

Lewis, D. J. (1979). Psychobiology of active and inactive memory. *Psychol. Bull.*, *86*(5), 1054–1083.

Loftus, E. F. (1997). Creating false memories. *Sci. Am.*, *277*(3), 70–75.

Loftus, E. F., & Yuille, J. C. (1984). Departures from reality in human perception and memory. In H. Weingartner & E. S. Parker (Eds.), *Memory consolidation: Psychobiology of cognition* (pp. 163–184). Hillsdale, NJ: Lawrence Erlbaum Associates.

Malinow, R., & Malenka, R. C. (2002). AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.*, *25*, 103–126.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.*, *24*, 897–931.

Martin, S. J., Grimwood, P. D., & Morris, R. G. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annu. Rev. Neurosci.*, *23*, 649–711.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and

failures of connectionist models of learning and memory. *Psychol. Rev.*, *102*(3), 419–457.
McGaugh, J. L. (1966). Time-dependent processes in memory storage. *Science*, *153*(742), 1351–1358.
McGaugh, J. L. (2000). Memory—A century of consolidation. *Science*, *287*(5451), 248–251.
McGaugh, J. L. (2004). Memory reconsolidation hypothesis revived but restrained: Theoretical comment on Biedenkapp and Rudy (2004). *Behav. Neurosci.*, *118*(5), 1140–1142.
McGaugh, J. L., & Krivanek, J. A. (1970). Strychnine effects on discrimination learning in mice: Effects of dose and time of administration. *Physiol. Behav.*, *5*(12), 1437–1442.
Milekic, M. H., & Alberini, C. M. (2002). Temporally graded requirement for protein synthesis following memory reactivation. *Neuron*, *36*(3) 521–525.
Miller, C. A., & Marshall, J. F. (2005). Molecular substrates for retrieval and reconsolidation of cocaine-associated contextual memory. *Neuron*, *47*(6), 873–884.
Miller, R. R., & Marlin, N. A. (1984). The physiology and semantics of consolidation: Of mice and men. In H. Weingartner & E. S. Parker (Eds.), *Memory consolidation: Psychobiology of cognition* (pp. 85–109). Hillsdale, NJ: Lawrence Erlbaum Associates.
Miller, R. R., & Springer, A. D. (1974). Implications of recovery from experimental amnesia. *Psychol. Rev. 81*(5), 470–473.
Millin, P. M., Moody, E. W., & Riccio, D. C. (2001). Interpretations of retrograde amnesia: Old problems redux. *Nat. Rev. Neurosci.*, *2*(1), 68–70.
Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science*, *160*(May), 203–204.
Müller, G. E., & Pilzecker, A. (1900). Experimentelle beitrage zur lehre vom gedachtnis. *Z. Psychol.*, Suppl. 1, 1.
Myers, K. M., & Davis, M. (2002). Systems-level reconsolidation: Reengagement of the hippocampus with memory reactivation. *Neuron*, *36*(3), 340–343.
Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: Computational, neuroimaging, and neuropsychological results. *Hippocampus*, *10*(4), 352–368.
Nader, K. (2003). Memory traces unbound. *Trends Neurosci.*, *26*(2), 65–72.
Nader, K., & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nat. Rev. Neurosci.*, *10*(3), 224–234.
Nader, K., Schafe, G. E., & LeDoux, J. E. (2000a). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726.
Nader, K., Schafe, G. E., & LeDoux, J. E. (2000b). The labile nature of consolidation theory. *Nat. Rev. Neurosci.*, *1*(3), 216–219.
Nader, K., & Wang, S. H. (2006). Fading in. *Learn. Memory*, *13*(5), 530–535.
Narayanan, R. T., Seidenbecher, T., Sangha, S., Stork, O., & Pape, H. C. (2007). Theta resynchronization during reconsolidation of remote contextual fear memory. *NeuroReport*, *18*(11), 1107–1111.
Pedreira, M. E., & Maldonado, H. (2003). Protein synthesis subserves reconsolidation or extinction depending on reminder duration. *Neuron*, *38*, 863–869.
Pedreira, M. E., Perez-Cuesta, L. M., & Maldonado, H. (2002). Reactivation and reconsolidation of long-term memory in the crab *Chasmagnathus*: Protein synthesis requirement and mediation by NMDA-type glutamatergic receptors. *J. Neurosci.*, *22*(18), 8305–8311.
Przybyslawski, J., & Sara, S. J. (1997). Reconsolidation of memory after its reactivation. *Behav. Brain Res.*, *84*(1–2), 241–246.
Quartermain, D., & McEwen, B. S. (1970). Temporal characteristics of amnesia induced by protein synthesis inhibitor: Determination by shock level. *Nature*, *228*(272), 677–678.
Ribot, T. (1881). *Les maladies de la memoire.* New York: Appleton-Century-Crofts.
Rodriguez, W. A., Rodriguez, S. B., Phillips, M. Y., & Martinez, J. L., Jr. (1993). Post-reactivation cocaine administration facilitates later acquisition of an avoidance response in rats. *Behav. Brain Res.*, *59*(1–2), 125–129.
Rose, J. K., & Rankin, C. H. (2006). Blocking memory reconsolidation reverses memory-associated changes in glutamate receptor expression. *J. Neurosci.*, *26*, 11582–11587.
Rudy, J. W., Biedenkapp, J. C., Moineau, J., & Bolding, K. (2006). Anisomycin and the reconsolidation hypothesis. *Learn. Memory*, *13*(1), 1–3.
Sangha, S., Scheibenstock, A., & Lukowiak, K. (2003). Reconsolidation of a long-term memory in *Lymnaea* requires new protein and RNA synthesis and the soma of right pedal dorsal 1. *J. Neurosci.*, *23*(22), 8034–8040.
Sangha, S., Scheibenstock, A., Morrow, R., & Lukowiak, K. (2003). Extinction requires new RNA and protein synthesis and the soma of the cell right pedal dorsal 1 in *Lymnaea stagnalis*. *J. Neurosci.*, *23*(30), 9842–9851.
Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *Am. Psychol.*, *54*(3), 182–203.
Schafe, G. E., & LeDoux, J. E. (2000). Memory consolidation of auditory Pavlovian fear conditioning requires protein synthesis and protein kinase A in the amygdala. *J. Neurosci.*, *20*(18), RC96.
Schafe, G. E., Nader, K., Blair, H. T., & LeDoux, J. E. (2001). Memory consolidation of Pavlovian fear conditioning: A cellular and molecular perspective. *Trends Neurosci., 24*(9), 540–546.
Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry*, *20*, 11–21.
Spear, N. (1973). Retrieval of memory in animals. *Psychol. Rev.*, *80*, 163–194.
Spear, N., & Mueller, C. (1984). Consolidation as a function of retrieval. In H. Weingarten & E. Parker (Eds.), *Memory consolidation: Psychobiology of cognition* (pp. 111–147). London: Laurence Erlbaum Associates.
Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.*, *99*(2), 195–231.
Squire, L. R. (2006). Lost forever or temporarily misplaced? The long debate about the nature of memory impairment. *Learn. Memory*, *13*(5), 522–529.
Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: A neurobiological perspective. *Curr. Opin. Neurobiol.*, *5*(2), 169–177.
Squire, L. R., & Barondes, S. H. (1972). Variable decay of memory and its recovery in cycloheximide-treated mice. *Proc. Natl. Acad. Sci. USA*, *69*(6), 1416–1420.
Squire, L. R., Slater, P. C., & Chace, P. M. (1976). Reactivation of recent or remote memory before electroconvulsive therapy does not produce retrograde amnesia. *Behav. Biol.*, *18*(3), 335–343.

Stollhoff, N., Menzel, R., & Eisenhardt, D. (2005). Spontaneous recovery from extinction depends on the reconsolidation of the acquisition memory in an appetitive learning paradigm in the honeybee (*Apis mellifera*). *J. Neurosci.*, *25*(18), 4485–4492.

Suzuki, A., Josselyn, S. A., Frankland, P. W., Masushige, S., Silva, A. J., & Kida, S. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *J. Neurosci.*, *24*(20), 4787–4795.

Tronson, N. C., Wiseman, S. L., Olausson, P., & Taylor, J. R. (2006). Bidirectional behavioral plasticity of memory reconsolidation depends on amygdalar protein kinase A. *Nat. Neurosci.*, *9*(2), 167–169.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annu. Rev. Psychol.*, *53*, 1–25.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.*, *80*(5), 359–380.

Valjent, E., Aubier, B., Corbille, A. G., Brami-Cherrier, K., Caboche, J., Topilko, P., et al. (2006). Plasticity-associated gene Krox24/Zif268 is required for long-lasting behavioral effects of cocaine. *J. Neurosci.*, *26*(18), 4956–4960.

Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, *425*(6958), 616–620.

Wang, S. H., Ostlund, S. B., Nader, K., & Balleine, B. W. (2005). Consolidation and reconsolidation of incentive learning in the amygdala. *J. Neurosci.*, *25*(4), 830–835.

Wiltgen, B. J., & Silva, A. J. (2007). Memory for context becomes less specific with time. *Learn. Memory*, *14*(4), 313–317.

Winocur, G., Moscovitch, M., & Sekeres, M. (2007). Memory consolidation or transformation: Context manipulation and hippocampal representations of memory. *Nat. Neurosci.*, *10*(5), 555–557.

48 The Dynamic Interplay between Cognitive Control and Memory

ELIZABETH A. RACE, BRICE A. KUHL, DAVID BADRE, AND ANTHONY D. WAGNER

ABSTRACT Cognitive control refers to the set of processes that guide thought and action in accordance with current goals. In this chapter we consider the manner in which cognitive control mechanisms guide mnemonic processing. First, we consider the architecture of prefrontal cortex (PFC) and review leading theories of how PFC operations support distinct forms of control. Next, we consider two illustrative and well-characterized situations in which PFC control guides mnemonic processing: (1) when competition between memories creates interference, and (2) when ineffective retrieval cues yield uncertainty. Finally, we consider the ways in which prior mnemonic experiences may reduce future interference and uncertainty, thereby easing the demands placed on PFC control mechanisms. Together, these considerations highlight the dynamic interplay between cognitive control and memory.

Cognitive control refers to the set of processes that guide thought and action in accordance with current goals. Central to higher cognitive function, cognitive control allows organisms to represent task demands, flexibly work with memory, and promote context- and goal-relevant information processing in the face of distraction. Control mechanisms are particularly important in unfamiliar situations or changing environments when acquired knowledge provides either insufficient or inappropriate information to satisfy current demands. The prefrontal cortex (PFC) is a fundamental component of the neural circuitry supporting cognitive control. By orchestrating the influence of past experience on present behavior, PFC mechanisms configure neural processing to optimize behavior.

In this chapter we explore the dynamic interaction between control mechanisms and memory, with a specific focus on prefrontal contributions to cognitive control. We begin with a brief description of the neural circuitry supporting cognitive control, focusing on the anatomy and connectivity of subregions within PFC. Next, we discuss current theories that characterize the mechanisms, functional organization, and regulation of cognitive control. Finally, we review functional neuroimaging and lesion evidence for the interaction between cognitive control and memory, with an emphasis on the interplay between mnemonic uncertainty, interference, and PFC-mediated control functions.

ELIZABETH A. RACE Neurosciences Program, Stanford University, Stanford, California
BRICE A. KUHL Department of Psychology, Stanford University, Stanford, California
DAVID BADRE Departments of Psychology and Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island
ANTHONY D. WAGNER Department of Psychology and Neurosciences Program, Stanford University, Stanford, California

PFC anatomy and connectivity

This chapter will focus on the function of four main subregions within PFC that have been implicated in cognitive control: ventrolateral, dorsolateral, frontopolar, and medial PFC (figure 48.1). Ventrolateral PFC (VLPFC) corresponds to the inferior frontal gyrus, encompassing pars orbitalis (area 47/12 in Petrides & Pandya, 2002), pars triangularis (~Brodmann's area [BA] 45), and pars opercularis (~BA 44). Following Badre and Wagner (2007), we refer to pars orbitalis as anterior VLPFC and pars triangularis as mid-VLPFC (note that these two regions have been collectively termed mid-VLPFC by Petrides & Pandya, 2002) and to pars opercularis as posterior VLPFC. Dorsolateral PFC (DLPFC) refers to regions within the middle frontal gyrus (areas 8, 9/46, and 46; Petrides & Pandya, 1999). In humans, the ventral bound of this region is defined by the inferior frontal sulcus and the dorsal bound by the superior frontal sulcus. Frontopolar cortex (~BA 10) corresponds to the most rostral portion of PFC, including portions of middle frontal gyrus. The medial wall of PFC includes portions of BAs 8, 9, and 10 and the anterior cingulate cortex (ACC; BAs 24 and 32). Though anatomically distinct, lateral and medial PFC subregions have been shown to be interconnected both with each other and with more posterior regions of cortex, including medial and lateral temporal cortex and posterior parietal cortex (Petrides & Pandya, 1999, 2002, 2007).

Theories of cognitive control

BIASED COMPETITION A prominent theory of cognitive control proposes that top-down signals derived from PFC *bias* processing in posterior brain regions in accordance with current task demands (e.g., Cohen & Servan-Schreiber, 1992; Desimone & Duncan, 1995; Miller & Cohen, 2001).

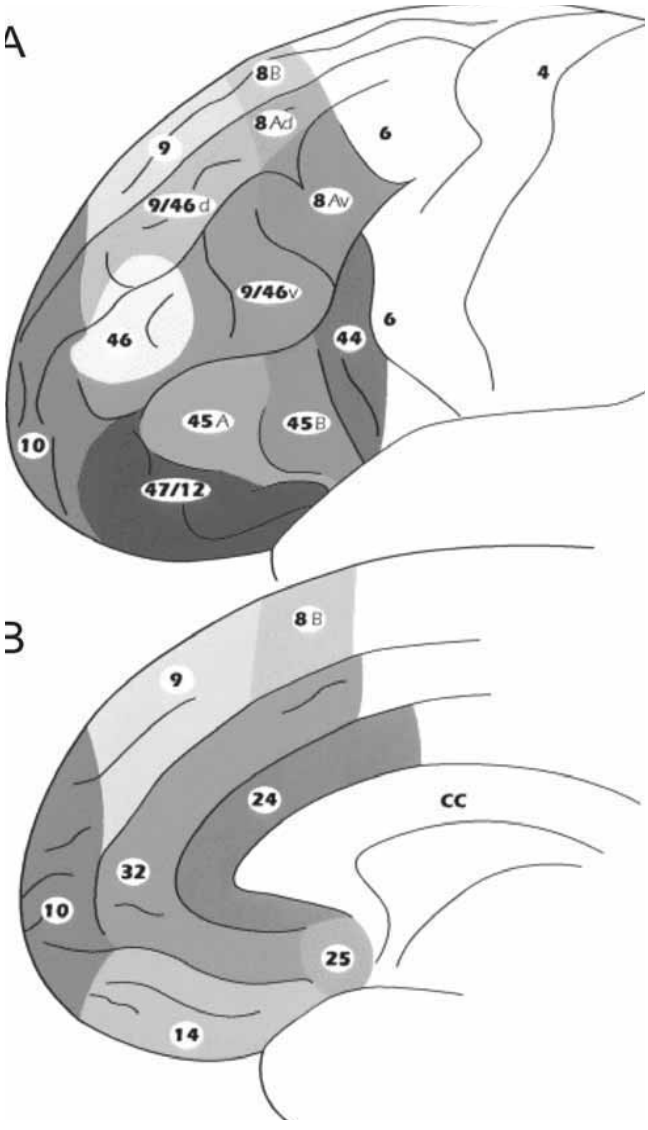


FIGURE 48.1 Anatomical subdivisions of the PFC. (*A*) Lateral view of left PFC depicting cytoarchitectonic areas (numbered). Anterior VLPFC corresponds to areas 47/12, mid-VLPFC corresponds to area 45, and posterior VLFPC corresponds to area 44. DLPFC corresponds to middle frontal gyrus including areas 8, 9/46, and 46. FPC corresponds to area 10. (*B*) Medial view of right PFC. ACC corresponds to areas 24 and 32. FPC corresponds to area 10. VLFPC, ventrolateral prefrontal cortex. DLPFC, dorsolateral prefrontal cortex. FPC, frontopolar cortex. ACC, anterior cingulate cortex. (Reprinted from M. Petrides & D. N. Pandya, 1999. Dorsolateral prefrontal cortex: Comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *Eur. J. Neurosci.*, *11*, 1011–1036. Copyright 1999, with permission from Blackwell Synergy.)

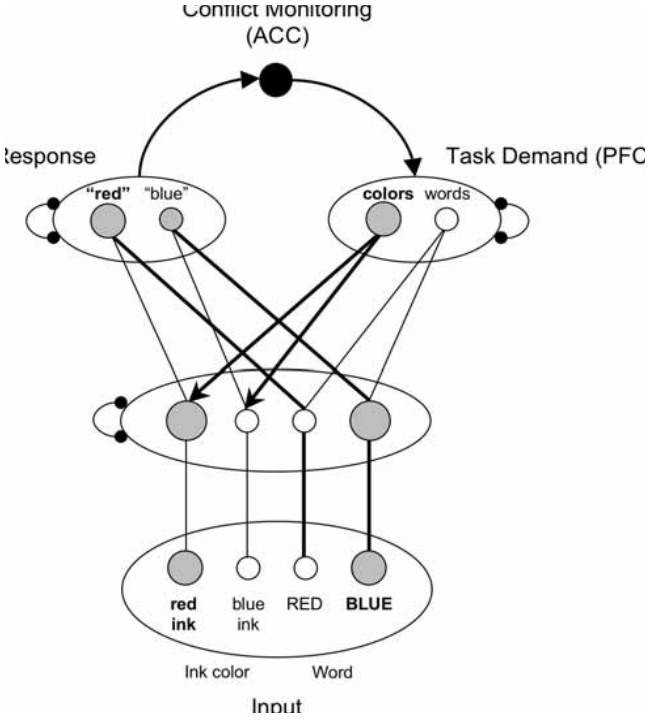


FIGURE 48.2 Model of PFC and anterior cingulate involvement during performance of the Stroop task. Circles represent processing units, which correspond to a population of neurons assumed to code a given piece of information. Lines represent connections between units, with heavier lines indicating stronger connections. Looped connections with black circles indicate mutual inhibition among units within that layer. In the Stroop task, subjects must name the ink color in which a word is presented, rather than read the word. The presentation of a conflict stimulus (the word "blue" displayed in red ink) activates (indicated by gray fill) input layer units representing "red ink" and the word "blue." The "colors" task demand unit is activated in PFC (gray fill), representing the current goal to name the color of the ink, and passes activation to the intermediate units in the color-naming pathway (indicated by arrows), increasing the activation of those units and biasing processing in favor of activity flowing along the color-naming pathway. This bias favors activation of the response unit ("red") corresponding to the color input (red ink), even though the connection weights in this pathway are weaker than in the word-reading pathway that would favor a response based on reading the word ("blue"). By computing the level of conflict (or the presence of simultaneously active representations in the response layer), ACC initially detects the need for this top-down bias from PFC. ACC, anterior cingulate cortex. (Adapted with permission from M. M. Botvinick, T. S. Braver, D. M. Barch, C. S. Carter, & J. D. Cohen, 2001. Conflict monitoring and cognitive control. *Psychol. Rev.*, *108*, 624–652. Copyright 2001, American Psychological Association.)

Specifically, the maintenance of task-relevant contextual representations in PFC has been proposed to bias establishment of appropriate mappings between sensory inputs, internal states, and motor outputs. In the absence of cognitive control, behavior is driven in an automatic, bottom-up fashion by representations that are most strongly activated by input cues. However, when weakly established (but task-relevant) representations must be selected in the face of competition from stronger (but task-irrelevant) representations, PFC control signals are thought to bias the flow of information processing to enhance the strength of the relevant representations and overcome the task-irrelevant competitors (Cohen, Dunbar, & McLelland, 1990).

Illustrative of this putative bias mechanism, consider the Stroop paradigm, wherein subjects are presented color words in different ink colors and are asked to name the ink color (figure 48.2). Presentation of a word strongly elicits the prepotent response to read the word, because subjects have more experience reading words than naming the color of word print. Thus, if the ink color is incongruent with the color word (e.g., "BLUE" in red ink), a prepotent response ("blue") must be overcome in favor of a weaker response ("red"). Biased competition theory proposes that lateral PFC represents the current task goal (e.g., name the ink color) and biases processing in color-naming pathways to favor the weaker but goal-relevant response (Cohen, Dunbar, & McClelland, 1990).

Importantly, top-down bias mechanisms have been argued to support a variety of functions, including working memory, selective attention, controlled retrieval from long-term memory, task switching, response inhibition, and response selection. While the biased competition theory proposes a central mechanism for cognitive control, there may be multiple types of control that differ in their form or domain. In the next sections, we describe several theories that focus on the functional architecture of control and its relationship to the organization of PFC.

The Dorsal-Ventral Hypothesis A complementary perspective on cognitive control suggests that dorsal and ventral regions of lateral PFC mediate dissociable, but interactive, forms of control (Petrides, 1994; Owen, Evans, & Petrides, 1996). In this view, control mechanisms supported by VLPFC and DLPFC operate over different loci or types of representations (Petrides, 1996). VLPFC mechanisms have been proposed to support controlled retrieval and selection of long-term knowledge stored in posterior cortices and the maintenance of these representations within working memory, while DLPFC mechanisms have been proposed to support the monitoring and manipulation of the representations retrieved and maintained by VLPFC (e.g., D'Esposito et al., 1998; Petrides, 2002).

Neuroimaging and lesion data support the proposal that DLPFC and VLPFC functionally differ. For example, lesions of mid-DLPFC (areas 9/46 and 46) produce impairments in the ability to order information in working memory (Petrides, 2000), and functional magnetic resonance imaging (fMRI) studies indicate that DLPFC activity increases during complex working memory tasks, such as when working memory loads are high (Rypma, Prabhakaran, Desmond, Glover, & Gabrieli, 1999), as well as when representations held in working memory must be reordered (D'Esposito, Postle, Ballard, & Lease, 1999; Postle, Berger, & D'Esposito, 1999; Wagner, Maril, Bjork, & Schacter, 2001) or updated (Salmon et al., 1996; Garavan, Ross, Li, & Stein, 2000). Similarly, within episodic retrieval tasks, DLPFC activation has often been associated with monitoring retrieved mnemonic information (e.g., Henson, Rugg, Shallice, & Dolan, 2000; Fletcher & Henson, 2001; Dobbins, Foley, Schacter, & Wagner, 2002; Rugg, Henson, & Robb, 2003; Achim & Lepage, 2005; Dobbins, Simons, & Schacter, 2004). In contrast, neuroimaging studies indicate that activity within VLPFC increases during the controlled retrieval and selection of information from long-term memory, as well as in the presence of mnemonic interference (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Jonides, Smith, Marschuetz, Koeppe, & Reuter-Lorenz, 1998; Bunge, Ochsner, Desmond, Glover, & Gabrieli, 2001; Badre & Wagner, 2002). We will further discuss VLPFC contributions to mnemonic control in the section on the interaction between control and memory.

Rostrocaudal Hierarchies In addition to apparent dorsal/ventral dissociations, accumulating evidence suggests that hierarchically organized cognitive control processes map to a functional gradient along the rostrocaudal axis of lateral frontal cortex (Christoff & Gabrieli, 2000; Fuster, 2001; Koechlin, Ody, & Kouneiher, 2003; Wood & Grafman, 2003; Bunge & Zelazo, 2006; Koechlin & Jubault, 2006; Petrides, 2006; Badre & D'Esposito, 2007; Botvinick, 2007; Koechlin & Summerfield, 2007; Badre, 2008). More caudal regions of frontal cortex, inclusive of premotor cortex, are thought to control processing at "lower" levels of representation in the stimulus-action processing hierarchy, such as response selection (figure 48.3). Progressively more anterior regions of frontal cortex are proposed to support control mechanisms that operate upon increasingly "higher" levels of representation (Christoff & Gabrieli, 2000; Badre, 2008), including more abstract higher-order plans or complex schemas. Functional organization along the horizontal axis of lateral PFC has also been characterized as mediating cross-temporal contingencies between past, present, and future events, with caudal PFC mechanisms guiding behavior based upon the immediate context in

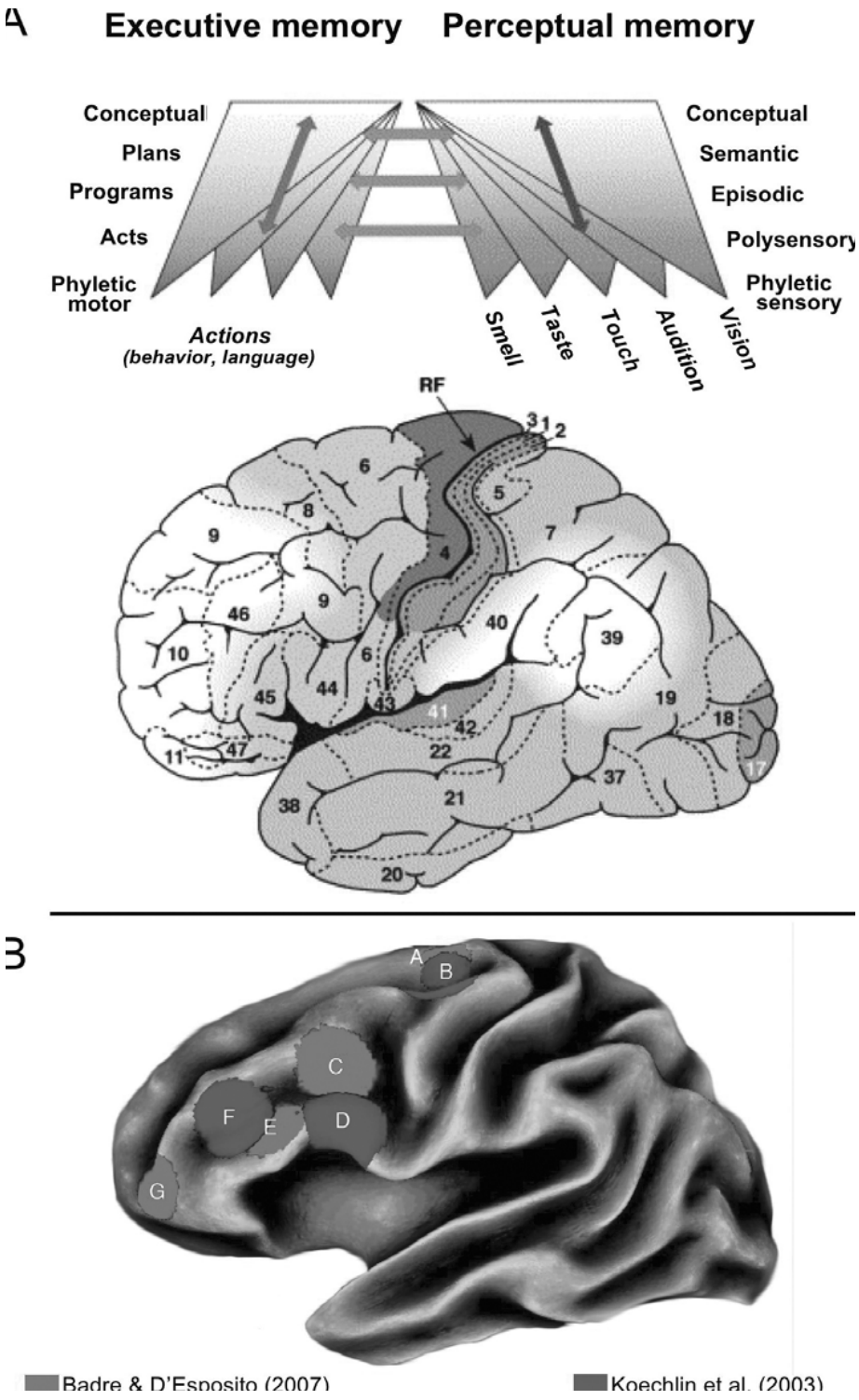

A
Executive memory
Perceptual memory
Conceptual
Plans
Programs
Acts
Phyletic motor
Actions
(behavior, language)
Conceptual
Semantic
Episodic
Polysensory
Phyletic sensory
Smell
Taste
Touch
Audition
Vision
RF
B
A
B
C
D
E
F
G
Badre & D'Esposito (2007)
Koechlin et al. (2003)

FIGURE 48.3 Hierarchical organization of cognitive representations in lateral cortex. (*A*) Schema of two hierarchies of cortical memory, executive memory, and perceptual memory, and the distribution of these hierarchies in frontal and posterior cortical regions, respectively. In frontal cortex, representations that are "higher" in the processing hierarchy are mapped to more rostral regions, and "lower"-level representations are mapped to more caudal regions. (Reprinted from J. M. Fuster, 2001, The prefrontal cortex—An update: Time is of the essence, *Neuron*, *30*, 319–333. Copyright 2001, with permission from Elsevier.) (*B*) Neuroimaging data providing evidence for representational hierarchies in frontal cortex. Spheres from Badre and D'Esposito (2007) (red) reflect foci of activation with experimental manipulations at different levels of representation: A, the response level; C, the feature level; E, the dimension level; G, the context level. Spheres from Koechlin, Ody, and Kouneiher (2003) (blue) reflect foci of activation with manipulations of different levels of control: B, sensory control; D, contextual control; F, episodic control. (Adapted with permission from D. Badre & M. D'Esposito, 2007, Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex, *J. Cogn. Neurosci.*, *19*, 2082–2099. Copyright 2007, with permission from the MIT Press.) (See color plate 60.)

which a stimulus occurs and more rostral PFC regions processing information that is successively more remote in time (Fuster, 2001; Braver, Reynolds, & Donaldson, 2003; Koechlin et al., 2003; Koechlin & Summerfield, 2007).

With its location at the most rostral extent of PFC, frontopolar cortex (FPC; ~BA 10; figure 48.1) may be positioned at the apex of the putative control hierarchy (Koechlin & Summerfield, 2007). While the precise functions of FPC remain to be determined, neuroimaging studies have consistently observed FPC activation during higher-level cognitive tasks, complex working memory tasks, and episodic retrieval (Fletcher & Henson, 2001; Ramnani & Owen, 2004). For example, FPC is recruited when previously selected goals or task-relevant information must be maintained in a pending state until ongoing subtasks are executed (Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999; Braver & Bongiolatti, 2002; Badre & Wagner, 2004; Koechlin & Hyafil, 2007). Similarly, FPC has been associated with higher-order functions such as integrating across multiple sources of information (Christoff et al., 2001; Bunge, Wendelken, Badre, & Wagner, 2004; Ramnani & Owen, 2004; Green, Fugelsang, Kraemer, Shamosh, & Dunbar, 2006; De Pisapia, Slomski, & Braver, 2007) or evaluating the products of internally generated information (Christoff, Ream, Geddes, & Gabrieli, 2003).

REGULATION OF CONTROL While control mechanisms supported by lateral PFC are thought to drive goal-relevant behavior, equally important are the mechanisms through which control is regulated. Substantial evidence indicates that regions within medial PFC, including the anterior cingulate cortex, serve this modulatory role (but see Fellows & Farah, 2005). Specifically, ACC computations have been alternately proposed to detect the presence of conflict (Botvinick, Cohen, & Carter, 2004; Kerns et al., 2004; MacDonald, Cohen, Stenger, & Carter, 2000), error likelihood (Brown & Braver, 2005), or uncertainty (Walton, Devlin, & Rushworth, 2004), and to signal lateral PFC mechanisms to increase top-down biasing of task-appropriate representations. For example, ACC may detect the presence of simultaneously active, competing representations (such as conflicting responses elicited by incongruent trials in the Stroop paradigm) and provide feedback signals to lateral PFC that up-regulate control (figure 48.2). Consistent with this proposal, imaging studies have documented functional coactivation of ACC and lateral PFC under situations of response and mnemonic conflict (e.g., Bunge, Burrows, & Wagner, 2004; Badre & Wagner, 2004; Kerns et al., 2004; Kuhl, Dudukovic, Kahn, & Wagner, 2007).

The basal ganglia (BG) have also been implicated in regulating PFC-mediated control processes. For example, in situations of response inhibition it has been argued that the subthalamic nucleus (a component of the BG) interacts with right VLPFC and preSMA such that initiated motor responses can be terminated (Aron & Poldrack, 2006; Aron et al., 2007). It has been argued, through computational models, that PFC-BG interactions also support cognitive operations, such as working memory performance (O'Reilly & Frank, 2006; Hazy, Frank, & O'Reilly, 2007). Specifically, this work has suggested that the BG gate PFC processing depending on task demands, with BG "learning" which PFC mechanisms to gate through dopamine-mediated reinforcement learning. This hypothesis has received support from recent evidence that PFC-BG interactions support working memory performance and that BG activation prior to the onset of working memory trials is predictive of the extent to which task-irrelevant information is successfully gated, or denied processing (McNab & Klingberg, 2008).

Interactions between control and memory

Having surveyed leading theories of how PFC implements cognitive control, we now consider the manner in which prefrontal control interacts with mnemonic operations. However, because there are numerous examples of such interactions across multiple forms and stages of memory and involving multiple PFC subregions (for reviews see Fletcher & Henson, 2001; Wagner, 2002; Buckner, 2003; Simons & Spiers, 2003), we restrict our focus to two examples of PFC involvement in mnemonic processing. Specifically, we consider how VLPFC mechanisms contribute to performance (1) when memory representations interfere with each other and (2) when ineffective retrieval cues yield uncertainty.

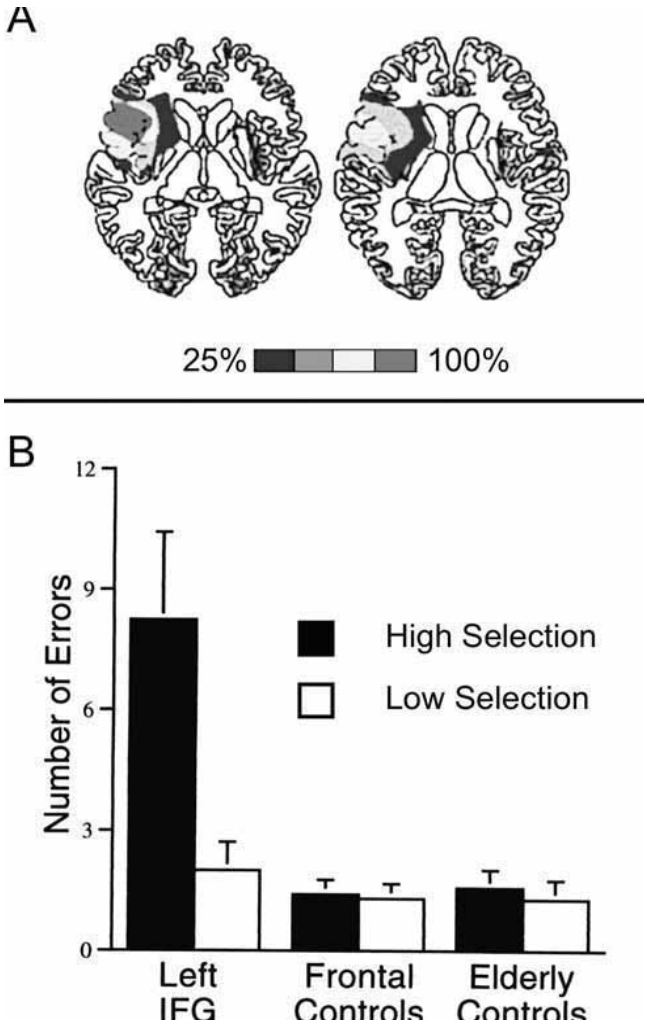
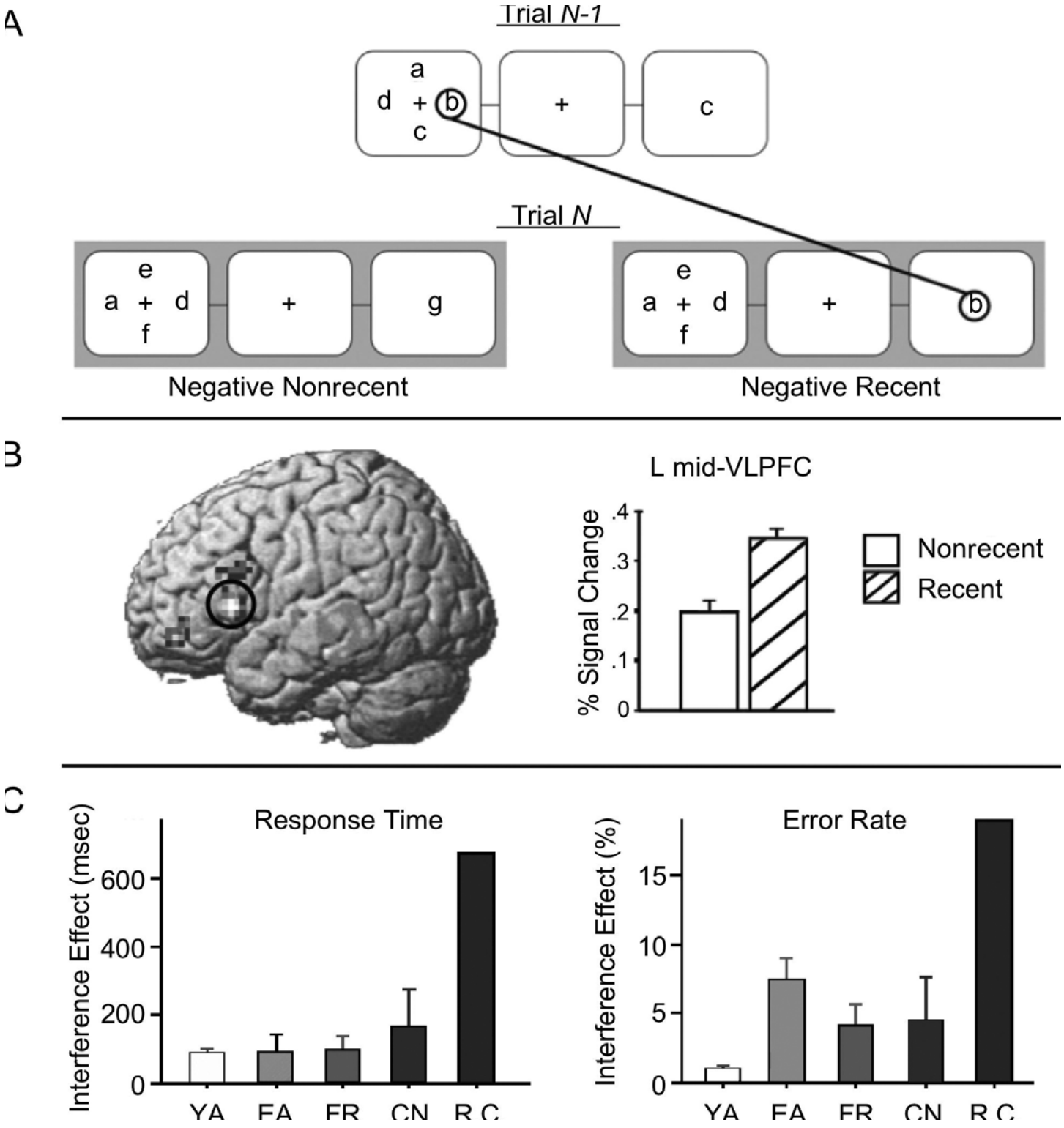


FIGURE 48.4 Damage to mid- and posterior VLPFC in humans impairs the ability to select relevant semantic information under high-selection demands. (*A*) Location of PFC lesions in patients with selection deficits on a verb-generation task. Scale represents amount of lesion overlap across patient group. (*B*) Mean number of errors in a task requiring subjects to generate semantically appropriate verbs for concrete nouns under high-selection demands (filled bars) versus low-selection demands (unfilled bars). Nouns in the high-selection group had a lower response-strength ratio (ratio of the relative response frequency of the most common completion to the relative response frequency of the second most common completion) than did nouns in the low-selection condition. Subject groups were composed of patients with lesions restricted to left inferior frontal gyrus (left IFG group), patients with frontal lesions outside of left IFG (frontal controls), and healthy older adults (elderly controls). (Adapted with permission from S. L. Thompson-Schill, D. Swick, M. J. Farah, M. D'Esposito, I. P. Kan, & R. T. Knight, 1998, Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings, *Proc. Natl. Acad. Sci. USA*, *95*, 15855–15860. Copyright 1999, with permission from National Academy of Sciences, U.S.A.)

INTERFERENCE Interference refers to the processing costs that arise when irrelevant representations compete with goal-relevant representations. For example, when making a trip to the grocery store to purchase a handful of items, one may find that remembering the items of interest becomes remarkably difficult while actually walking down the grocery store aisles, owing to the salience of countless products that are *not* the items of interest. Overcoming interference requires a mechanism that selects relevant representations from the set of all active representations. Evidence accumulated across semantic memory, working memory, and episodic memory paradigms has led to the hypothesis that left mid-VLPFC, in particular, supports a *selection* mechanism that plays a fundamental role in resolving mnemonic interference (for review see Badre & Wagner, 2007).

The selection hypothesis of VLPFC function was originally formulated within the context of semantic retrieval. In a seminal paper, Thompson-Schill and colleagues demonstrated that left mid- and posterior VLPFC are engaged to

FIGURE 48.5 Evidence for left mid-VLPFC involvement in resolving interference during the Sternberg working memory task. (*A*) The interference variant of the Sternberg working memory paradigm in which subjects maintain a set of four letters in working memory until a probe letter appears, at which point they indicate whether the probe is a member of the currently maintained set (positive probe) or is not a member of the currently relevant set (negative probe). Interference occurs when a negative probe is a member of the immediately preceding set (negative recent) relative to a negative probe that is not a member of the immediately preceding set (negative nonrecent). (*B*) Greater activation in left mid-VLPFC (circled), as measured by fMRI, occurs during negative recent (hatched bar) compared to negative nonrecent trials (unfilled bar), reflecting greater recruitment of left mid-VLPFC in the presence of interference in working memory. (*A* and *B* adapted from D. Badre & A. D. Wagner, 2005, Frontal lobe mechanisms that resolve proactive interference, *Cereb. Cortex*, *15*, 2003–2012. Copyright 2005, with permission from Oxford University Press.) (*C*) Damage to left middle and inferior frontal gyri in patient R.C. impairs the ability to successfully reject negative recent probes. Patient R.C., a 51-year-old male with a significant lesion in left middle and inferior frontal gyri, showed a pronounced interference effect in both response times (left panel) and accuracy (right panel) compared to four control groups: control subjects that were matched in age and education to R.C. (Controls: *CN*); frontal patients with damage outside of left mid-VLPFC (Frontal Patients: *FR*); older adults matched in age and education to the frontal patient group (Elderly: *EA*); and a group of young adults (Young: *YA*). (Adapted with permission from S. L. Thompson-Schill, J. Jonides, C. Marshuetz, E. E. Smith, M. D'Esposito, I. P. Kan, R. T. Knight, & D. Swick, 2002, Effects of frontal lobe damage on interference effects in working memory, *Cogn. Affective Behav. Neurosci.*, *2*, 109–120. Copyright 2002, with permission from Psychonomic Society, Inc.)


A
Trial N-1
a
d + b
c
+
c
Trial N
e
a + d
f
+
g
Negative Nonrecent
e
a + d
f
+
b
Negative Recent
B
L mid-VLPFC
% Signal Change
.4
.3
.2
.1
0
Nonrecent
Recent
C
Response Time
Interference Effect (msec)
600
400
200
0
YA
FA
FR
CN
R C
Error Rate
Interference Effect (%)
15
10
5
0
YA
FA
FR
CN
R C

the extent that semantic decisions require selecting goal-relevant information in the face of competition (Thompson-Schill et al., 1997). For example, in one task subjects were shown nouns and required to generate semantically related verbs; critically, some of the nouns were associated with a dominant verb (e.g., "scissors" strongly elicits "cut"; a low-selection situation), whereas other nouns were associated with multiple verbs (e.g., "wheel" may elicit "turn," "steer," and "drive"; a high-selection situation). Functional MRI revealed greater left mid- and posterior VLPFC activation during generation under high- relative to low-selection demands (for related findings, see Thompson-Schill, D'Esposito, & Kan, 1999; Badre, Poldrack, Paré-Blagoev, Insler, & Wagner, 2005). Subsequent work demonstrated that damage to left mid- and posterior VLPFC in humans impairs the ability to select relevant semantic representations—specifically when competition is present—establishing the necessity of this region for resolving semantic interference (figure 48.4; Thompson-Schill et al., 1998).

Additional evidence for the role of left mid-VLPFC in resolving interference comes from studies using the interference variant of the Sternberg working memory paradigm (figure 48.5). In this paradigm, each trial requires the encoding and maintenance of a set of stimuli in working memory and determination of whether a subsequently presented test probe is or is not a member of the currently maintained set (trial *N*). Interference occurs when the test probe is not a member of the currently maintained set but was a member of the previously maintained set (trial $N - 1$)—"negative recent" probes. The now classic finding is that "negative recent" probes elicit greater activation in left mid-VLPFC than do "negative nonrecent" probes—trials requiring the same decision but without interference (figure 48.5; e.g., Jonides et al., 1998; D'Esposito, Postle, Jonides, & Smith, 1999; Bunge et al., 2001; Badre & Wagner, 2005; Nee, Jonides, & Berman, 2007). Moreover, the ability to successfully reject negative recent probes is compromised by left mid-VLPFC damage (Thompson-Schill et al., 2002; figure 48.5) or disruption by means of transcranial magnetic stimulation (Feredoes, Tononi, & Postle, 2006). Mechanistically, it has been argued that rejecting negative recent probes engages left mid-VLPFC because accurate task performance requires identifying (selecting) the relevant context for the familiar negative probe (i.e., that it appeared in the last trial) so that it can be appropriately rejected (Badre & Wagner, 2005; for alternative interpretations, see Jonides & Nee, 2006).

Within the domain of episodic memory, selection to overcome interference likely plays a role during both encoding and retrieval. In a classic PET study, Dolan and Fletcher (1997) measured neural responses during the encoding of word pairs, manipulating the extent to which prior learning interfered with current encoding (i.e., *proactive interference*). Specifically, subjects first studied a list of word pairs (e.g., "DOG-BOXER"); next, a second list of word pairs was studied, containing either repeated pairs, completely novel pairs, or pairs that partially overlapped with previously studied pairs (e.g., "SPORTSMAN-BOXER"—the proactive interference condition). Dolan and Fletcher observed that left lateral PFC, inclusive of left mid-VLPFC, was highly sensitive to the presence of interference, as this region was differentially engaged when subjects were encoding word pairs that overlapped with previously studied pairs. Additional findings relating left mid-VLPFC to the resolution of proactive interference have been reported in more recent fMRI studies of episodic encoding (Fletcher, Shallice, & Dolan, 2000; Henson, Shallice, Josephs, & Dolan, 2002), complementing neuropsychological observations that damage to lateral PFC results in an increased susceptibility to proactive interference (e.g., Shimamura, Jurica, Mangels, Gershberg, & Knight, 1995; Smith, Leonard, Crane, & Milner, 1995). It has been argued that, during episodic encoding, left mid-VLPFC-mediated selection may allow for relevant semantic associations between word pairs to be favored in the face of interference from previously learned, irrelevant associations (Henson et al., 2002).

Left mid-VLPFC engagement has also been observed during episodic retrieval situations that are well characterized as requiring selection. For example, with an increase in the number of competing associates that interfere with retrieval of a target associate, left lateral PFC, inclusive of left mid-VLPFC, displays a corresponding increase in retrieval-related activation (Sohn, Goode, Stenger, Carter, & Anderson, 2003; Sohn et al., 2005; Danker, Gunn, & Anderson, 2008). Likewise, when a retrieval task involves recollecting a specific detail of an encoding event over other possible event details (e.g., as in source memory tasks), left mid-VLPFC, among other regions, is engaged (e.g., Nolde, Johnson, & D'Esposito, 1998; Dobbins et al., 2002; Cabeza, Locantore, & Anderson, 2003; Dobbins & Wagner, 2005; Lundstrom, Ingvar, & Petersson, 2005). Importantly, left mid-VLPFC is distinguished from other lateral PFC regions engaged during source retrieval in that it supports source recollection in a domain-general manner (Dobbins & Wagner, 2005). These data complement neuropsychological observations that patients with lateral PFC damage are particularly impaired at attributing retrieved information to its relevant source (Janowsky, Shimamura, & Squire, 1989).

In summary, extant data provide strong support for the hypothesis that left mid-VLPFC mediates the resolution of interference by selecting goal-relevant representations in the face of competition from irrelevant representations. While we have focused on the role of selection in working-memory, semantic-retrieval, and episodic-memory paradigms, it is worth noting that mid-VLPFC selection has also been asso-

ciated with overcoming proactive interference during task switching (Badre & Wagner, 2006). As such, this selection mechanism does not appear to support retrieval, per se (Thompson-Schill et al., 1997); rather, selection likely operates postretrieval such that goal-relevant representations can be favored over goal-irrelevant representations (Badre & Wagner, 2007).

Uncertainty While left mid-VLPFC (~BA 45) is thought to support selection between activated representations, a central question is whether there are additional PFC mechanisms that support the top-down activation of representations under other situations of uncertainty. Here we define uncertainty as the situation in which goal-relevant representations are not automatically activated because of ineffective triggering cues. Under such situations, strategic activation, or *controlled retrieval*, of goal-relevant representations is required to recover relevant knowledge (Wagner, Paré-Blagoev, Clark, & Poldrack, 2001; Badre & Wagner, 2002; Badre et al., 2005; Badre & Wagner, 2007). Extant data indicate that anterior VLPFC (area 47/12) mediates controlled retrieval, with the left homologue differentially supporting such retrieval from semantic memory and the right homologue from visual associative memory.

Evidence for the distinction between selection and controlled retrieval comes from an fMRI study that varied demands on each of these putative control processes (Badre et al., 2005). In that study, controlled retrieval demands were manipulated by varying the strength of the semantic association between a cue and target in a task in which subjects were required to identify semantic associates (targets) of particular cues. For example, identifying the semantic relationship between strongly associated nouns such as "candle" and "flame" places low demands on controlled retrieval, relative to weakly associated nouns such as "candle" and "halo." The difference in controlled retrieval demands is due to the fact that "candle" is more likely to generate bottom-up activation of the associated concept "flame," thereby facilitating identification of a semantic relationship; "candle," however, is less likely to elicit bottom-up activation of weakly associated concepts such as "halo," meaning that identification of a semantic relationship between these stimuli requires top-down semantic search. Within this same decision task, selection demands were independently manipulated by varying the extent to which irrelevant semantic information was likely to interfere (e.g., by including distracters that were either strongly or weakly interfering). Consistent with the selection literature, Badre and colleagues (2005) reported increases in left mid-VLPFC activity as selection demands increased (figure 48.6). In contrast, increases in controlled retrieval demands were associated with increased engagement of left anterior VLPFC and middle temporal cortex—regions that were not modulated by selection (see also Wagner, Paré-Blagoev, et al., 2001). The coactivation of left anterior VLPFC and middle temporal cortex suggests a frontal-temporal interaction in which left anterior VLPFC provides a top-down bias that activates semantic representations stored in temporal cortex.

Functional dissociations between left mid-VLPFC and left anterior VLPFC have also been observed in the context of short-term semantic priming (Gold et al., 2006) and episodic retrieval (Danker et al., 2008). For example, Gold and colleagues (2006) used a lexical decision priming task to identify regions in which neural processing demands were (1) decreased with the presentation of semantically related primes and (2) increased with the presentation of semantically unrelated (interfering) primes. These two situations provide a compelling parallel between the controlled retrieval and selection distinction explored by Badre and colleagues (2005). For example, when a "related" semantic prime is presented (e.g., "SPOON" as a prime for the target "FORK"), the prime should elicit bottom-up semantic activation that reduces the demand for *controlled retrieval* once the target appears (i.e., the prime has already activated the relevant semantic information). On the other hand, "unrelated" semantic primes (e.g., "SPOON" as a prime for "COAT") elicit activation of irrelevant semantic information that may interfere with access to target-related information, thus requiring subsequent *selection* of relevant target-related information in the face of irrelevant information. Strikingly, the presentation of "related" primes resulted in reduced engagement of left anterior VLPFC and middle temporal cortex, presumably because of reduced controlled retrieval demands, relative to a neutral prime control condition. In contrast, the increased selection demands associated with "unrelated" primes resulted in increased engagement of left mid-VLPFC, relative to the neutral prime condition. Paralleling these findings, Danker, Gunn, and Anderson (2008) observed that left mid-VLPFC and anterior VLPFC functionally dissociate during episodic retrieval, with the former being sensitive to mnemonic competition (fan size) and associative memory strength and the latter being selectively sensitive to associative memory strength.

Together, these studies of semantic retrieval (Badre et al., 2005; Gold et al., 2006) and episodic retrieval (Danker, Gunn, & Anderson, 2008; see also Dobbins & Wagner, 2005) provide compelling evidence for a dissociation between a selection mechanism supported by left mid-VLPFC and a controlled retrieval mechanism supported by left anterior VLPFC that interacts with middle temporal cortex. The argument that left anterior VLPFC, in particular, supports controlled semantic retrieval is also supported by evidence that neural disruption (by means of transcranial magnetic stimulation) of left anterior VLPFC, but not left posterior VLPFC, interferes with semantic—but not phonological—processing (Gough, Nobre, & Devlin, 2005).

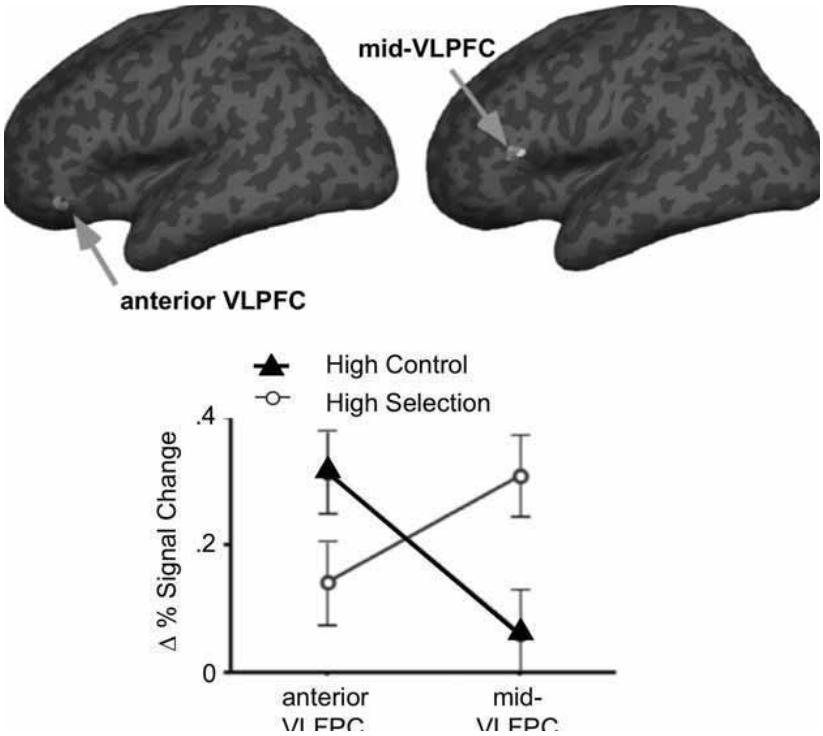


FIGURE 48.6 Left ventrolateral PFC is differentially engaged during controlled retrieval and selection from semantic memory. During a semantic decision task, participants were presented with target words beneath a cue word. On each trial, participants determined which of the target words was semantically related to the cue. Selection demands were manipulated by varying the task requirements for each trial (either a global relatedness judgment or a more specific feature similarity judgment that entailed higher selection demands) and by varying the extent to which irrelevant semantic information was likely to interfere with the decision (i. e., the distracter could be a preexperimental associate of the cue, but not along the relevant dimension). Controlled retrieval demands were manipulated by varying the strength of the association between a cue and the correct target. Greater controlled retrieval is necessary under conditions of weak cue-target associative strength because of diminished bottom-up activation of relevant knowledge. The top panel shows the location of anterior VLPFC and mid-VLPFC regions of interest. The fMRI data from these regions of interest (bottom panel) reveal a crossover interaction wherein anterior VLPFC displays greater activity with high control demands than with high selection demands, and mid-VLPFC displays greater activity with high selection demands than with high control demands. (Adapted from D. Badre & A. D. Wagner, 2007, Left ventrolateral prefrontal cortex and the cognitive control of memory, *Neuropsychologia*, *45*, 2883–2901. Copyright 2007, with permission from Elsevier.)

It should be noted, however, that controlled retrieval does not render selection unnecessary. That is, the combination of automatic and controlled semantic retrieval may result in the activation of multiple representations, from which a subset must be selected. Indeed, Badre and colleagues (2005) describe conditions in which both selection and controlled retrieval demands were high, and these situations engaged both left anterior VLPFC and left mid-VLPFC (see also Danker et al.). Thus, while distinct VLPFC subregions appear to support dissociable forms of cognitive control, these functionally separable regions may act in concert when automatic retrieval is insufficient to arrive at mnemonic goals (Kostopoulos & Petrides, 2003, 2008). Moreover, given the dorsal-ventral hypothesis of prefrontal contributions to cognitive control, it is worth noting that PFC correlates of controlled retrieval and selection have been concentrated in VLPFC, rather than DLPFC, subregions.

Decreased PFC demands through mnemonic suppression and prediction

Thus far, we have described how the recruitment of PFC control processes facilitates achievement of current mnemonic goals. In this final section, we consider how past experience can favor goal-appropriate representations and reduce future demands on cognitive control. We describe evidence for modulation of control by (1) prior acts of selection that strengthen relevant memories and weaken interfering memories, and (2) experience-dependent plasticity that strengthens memory-based predictions to reduce uncertainty at multiple levels of processing between stimulus input and response output.

REDUCED INTERFERENCE Although the presence of competition during retrieval may require PFC mechanisms

that implement interference resolution (e.g., Thompson-Schill et al., 1997; Sohn et al., 2003; Dobbins & Wagner, 2005; Sohn et al., 2005), demands on PFC control mechanisms often change with experience. For example, memories that are repeatedly selected during retrieval accrue a competitive advantage over other memories that are selected against. This advantage stems from both the strengthening of selected memories (e.g., Roediger & Karpicke, 2006) and the weakening of interfering, selected-against memories (M. Anderson, 2003). These adaptive changes in memory strength are thought to "benefit" future processing by favoring memories that are likely to be relevant in the future (J. Anderson, 2007) and reducing interference from memories that are likely to remain irrelevant. Indeed, general support for the processing benefits associated with prior acts of selection comes from fMRI observations of reduced lateral PFC engagement across repeated acts of episodic retrieval relative to initial acts (e.g., Henson et al., 2002; Law et al., 2005). Moreover, electrophysiological evidence indicates that the engagement of PFC during initial selective retrieval is predictive of later forgetting (weakening) of interfering memories, suggesting that reductions in interference occur as a result of prior PFC-mediated mnemonic selection (Johansson, Aslan, Bäuml, Gabel, & Mecklinger, 2007).

Building on these observations, a recent fMRI study examined whether the PFC control mechanisms that support initial mnemonic selection also "benefit"—in terms of reduced subsequent processing demands—from the weakening of interfering memories (Kuhl et al., 2007). At a behavioral level, Kuhl and colleagues (2007) observed that repeated selective retrieval of target memories elicits forgetting of interfering memories, replicating prior observations of retrieval-induced forgetting (M. Anderson, Bjork, & Bjork, 1994; Levy & Anderson, 2002). Critically, when this behavioral effect was related to functional activation during the repeated acts of selective retrieval, the data revealed that the extent to which interfering memories were forgotten was tightly correlated with PFC processing benefits that occurred across the repeated acts of selective retrieval. Specifically, ACC and right anterior VLPFC displayed robust decreases in engagement during future target memory remembering to the extent that interfering memories were forgotten (figure 48.7).

While Kuhl and colleagues' (2007) data reveal the neural processing benefits of mnemonic filtering (for related findings, see M. Anderson et al., 2004; Depue, Curran, & Banich, 2007), it is important to emphasize that these benefits are obtained only when one's memory goals remain constant. By contrast, when previously interfering and selected-against memories later become goal-relevant, the weakening that these memories suffered results in *increased* demands on ACC and right anterior VLPFC processes during their subsequent retrieval (Kuhl, Kahn, Dudukovic, & Wagner, 2008). This dynamic interplay between cognitive control and memory highlights how experience-dependent changes in memory strength and mnemonic competition yield cognitive control benefits and costs, as evidenced by decreasing and increasing demands on PFC control mechanisms during future acts of remembering.

Reduced Uncertainty Experience-dependent learning also reduces demands on PFC-mediated control by decreasing uncertainty associated with previously encountered stimuli. Illustrative of this point is the phenomenon of *repetition priming*, a form of nondeclarative (or implicit) memory that is expressed behaviorally as faster reaction times, increased response accuracy, or otherwise biased responding when stimuli are repeatedly processed (Tulving & Schacter, 1990; Roediger & McDermott, 1993). For example, stimulus classification decisions—for example, "Is a horse animate?"—are speeded with repetition, reflecting the behavioral benefits of previous stimulus processing. At the neural level, cortical regions that are active during initial stimulus processing frequently show reduced responses during subsequent stimulus processing (e.g., Raichle et al., 1994; Gabrieli et al., 1996; Schacter & Buckner, 1998; Wiggs & Martin, 1998; Henson, 2003)—a phenomenon that has been referred to as *repetition suppression*, *neural priming*, or *fMRI adaptation*. For example, stimulus repetition in the visual domain is associated with reduced activation in visual cortical areas, as expressed in reduced neural firing rates (Desimone, 1996) and reduced PET/fMRI activation (Wiggs & Martin, 1998; Wagner & Koutstaal, 2002). These neural activation reductions are generally thought to reflect computational savings or more efficient processing in neural networks supporting stimulus perception.

While perceptual priming facilitates processing in sensory cortical regions, other forms of priming are associated with repetition suppression in lateral PFC. In particular, *conceptual priming*—implicit memory at the level of semantic or conceptual information—is typically associated with activation reductions in left-lateralized frontotemporal regions (figure 48.8*A*), including left VLPFC and regions within inferior and lateral temporal cortex (Demb et al., 1995; Wagner, Desmond, Demb, Glover, & Gabrieli, 1997; Buckner et al., 1998; Gabrieli et al., 1996). Conceptual priming is dissociable from perceptual priming in that conceptual priming is invariant to changes in perceptual input across repetitions (e.g., priming will occur across stimulus modality changes such as auditory to visual), whereas perceptual priming occurs to the extent that there is perceptual overlap across repetitions (e.g., words appearing in the same font or same modality) (Roediger & McDermott,

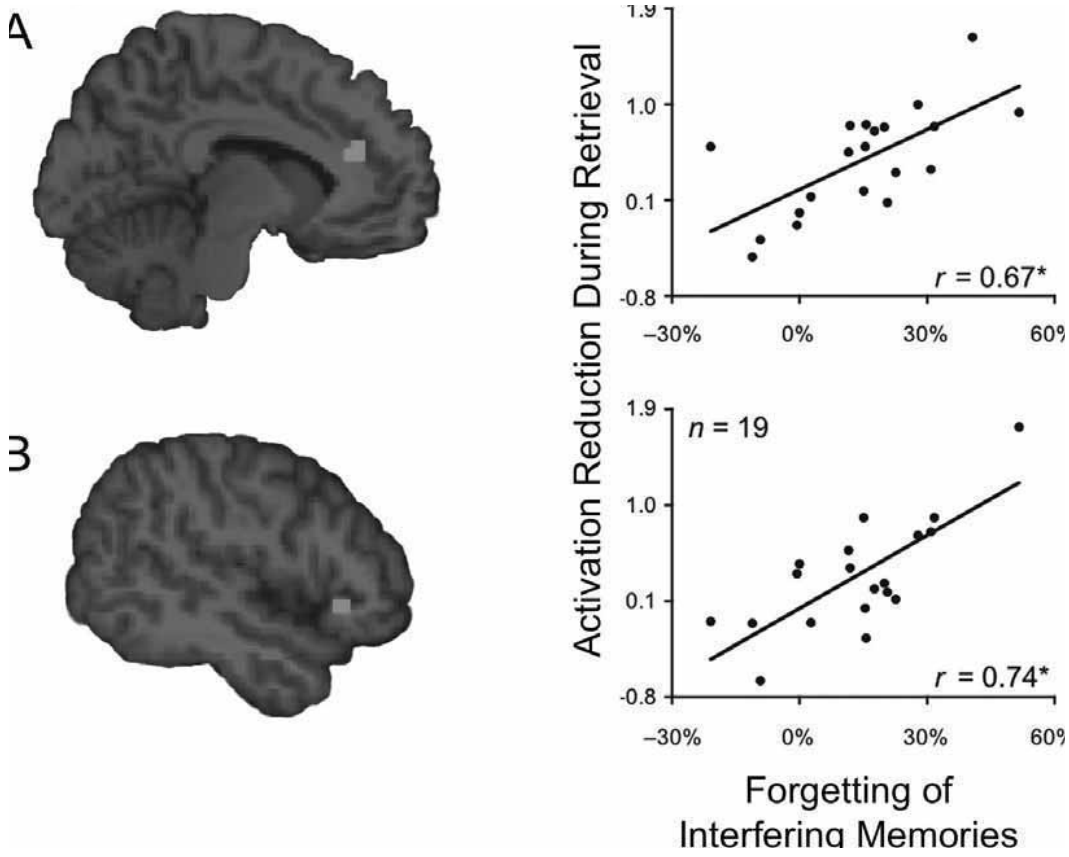


FIGURE 48.7 Demands on PFC control mechanisms that support initial mnemonic selection are reduced with the weakening of interfering memories. During repeated, selective retrieval of goal-relevant memories, fMRI activation reductions in (*A*) ACC and (*B*) right anterior VLPFC were correlated with the behavioral evidence that interfering memories were later forgotten, suggesting that processing demands on these regions were reduced to the extent that irrelevant memories were forgotten. (Adapted from B. A. Kuhl, N. M. Dudukovic, I. Kahn, & A. D. Wagner, 2007, Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nat. Neurosci.*, *10*, 908–914. Copyright 2007, reprinted by permission from Macmillan Publishers, Ltd.)

1993; Badgaiyan, Schacter, & Alpert, 2001; Carlesimo et al., 2003).

Although the repetition suppression in left VLPFC and lateral temporal cortex that accompanies conceptual priming is consistent with the hypothesis that these regions interact during controlled retrieval of semantic information (Badre et al., 2005; Gold et al., 2006), at present there is debate regarding the processes underlying repetition suppression in these cortical areas. The dominant, or traditional, view is that representations in cortical regions that store conceptual information are "tuned" with experience, such that previously accessed information is more effectively activated during future processing (Wiggs & Martin, 1998; Grill-Spector, Henson, & Martin, 2006; figure 48.8). Several mechanisms have been proposed to support such cortical "tuning" within a population of neurons, including reductions in overall activation (fatigue model), a reduction in the number of responsive neurons (sharpening model), and faster processing or settling time (facilitation model). Viewed in this light, left VLPFC reductions in conceptual priming tasks may reflect reduced control demands owing to increased availability of item-related knowledge.

By contrast, an alternative—though not mutually exclusive—account of repetition suppression in VLPFC is that prior processing of a stimulus results in "stimulus-response" learning that facilitates subsequent mappings between the stimulus and a decision or response (Dobbins, Schnyer, Verfaellie, & Schacter, 2004; Schacter, Dobbins, & Schnyer, 2004). For example, when repeatedly asked, "Is a horse animate?" subsequent performance can be facilitated by direct retrieval of a learned association between the "stimulus" with the relevant "response" ("yes"). Thus, while the retrieval of response information previously associated with a stimulus does not reflect facilitated conceptual processing (rather, it may enable the bypassing of controlled semantic retrieval), "stimulus-response" learning may nonetheless reduce demands on PFC control mechanisms that support decision or response selection (Schacter et al., 2004; Schacter, Wig, & Stevens, 2007).

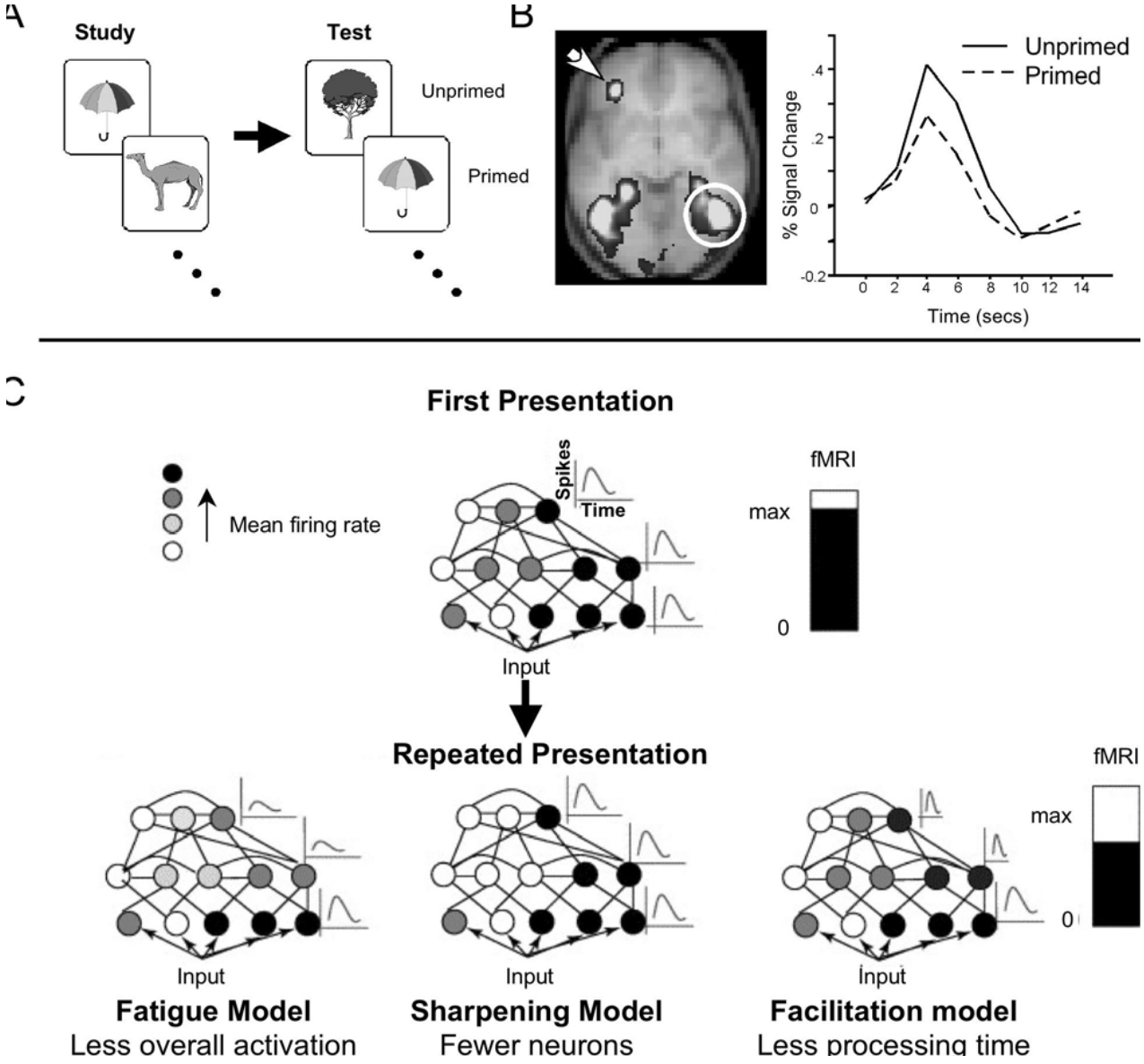


FIGURE 48.8 Repetition priming paradigm, neural priming effects, and hypothesized mechanisms of "cortical tuning." (*A*) In semantic priming tasks, subjects initially study a set of stimuli (e.g., pictures or words), making a semantic decision (e.g., size judgment) about those stimuli. Subsequently, during the critical test phase, semantic decisions are made about previously studied (primed) and novel (unprimed) stimuli. Typically, improved behavioral performance measures (e.g., reaction times and accuracy) are observed for primed compared to unprimed stimuli. (*B*) Functional MRI scanning during the test phase of a semantic classification priming task revealed activation reductions in fusiform (circled) and left ventrolateral PFC (arrow) for primed compared to unprimed stimuli. (Data from W. Koutstaal, A. D. Wagner, M. Rotte, A. Maril, R. L. Buckner, & D. L. Schacter, 2001, Perceptual specificity in visual object priming: Functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex, *Neuropsychologia*, 39, 184–199. Copyright 2001, with permission from Elsevier.) (*C*) Proposed experience-dependent changes in a neural network representing visual object features. First presentation of a stimulus activates a network of neurons (circles) coding for relevant and irrelevant features of the stimulus. Repeated presentation "tunes" the stimulus representation, reducing the overall firing rate across this network as well as the associated fMRI signal. Possible mechanisms supporting cortical "tuning" in a population of neurons with repeated stimulus presentation include less overall activation (fatigue model), a reduction in the number of responsive neurons (sharpening model), and faster processing or settling time (facilitation model). (Adapted with permission from K. Grill-Spector, R. Henson, & A. Martin, 2006, Repetition and the brain: Neural models of stimulus-specific effects, *Trends Cogn. Sci.*, *10*, 14–23. Copyright 2006, with permission from Elsevier.)

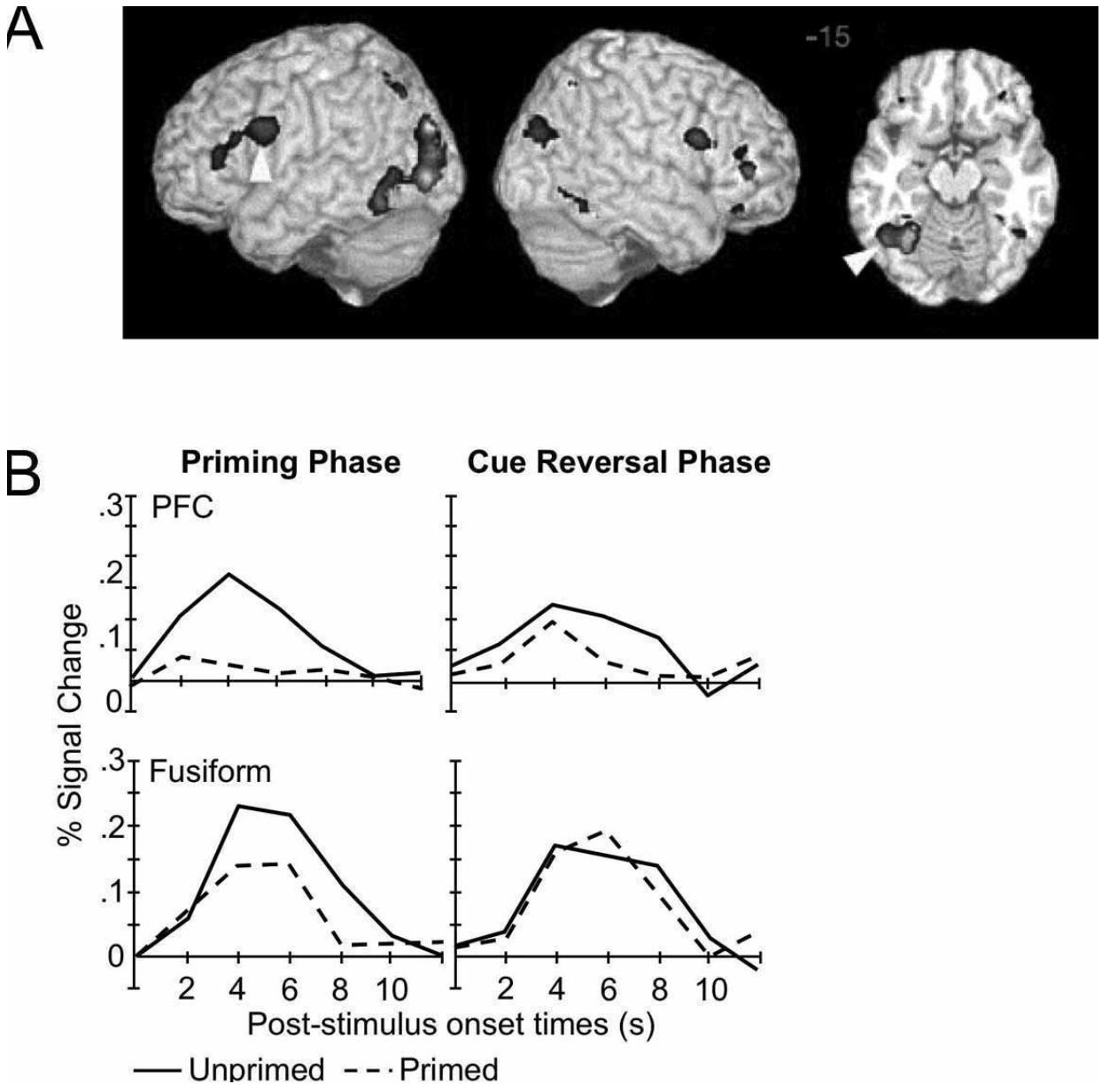


FIGURE 48.9 Contributions of "response learning" to neural priming during a semantic classification task. Subjects semantically classified visually presented objects ("Bigger than a shoe box?") that were presented once (unprimed) or three times (primed) and responded with a yes/no response. During a subsequent cue reversal phase the task cue was inverted ("Smaller than a shoe box?"), and half of the items from the previous priming phase were represented along with a new set of unprimed items. (*A*) Functional MRI scanning revealed regions displaying reductions in the neural priming signal (difference in activation between primed and unprimed trials) in the cue reversal relative to the priming phase (left panel arrow points to left posterior VLPFC [BA 9/44]; right panel arrow points to left fusiform [BA 37]). (*B*) Hemodynamic time courses from the two regions of interest indicated in *A*. Both posterior VLPFC and fusiform cortex showed significant neural priming during the priming phase when the classification rule was held constant. Inversion of the classification rule in the cue inversion phase reduced neural priming in posterior VLPFC and eliminated priming in fusiform cortex. The disruption of neural priming in the cue reversal phase suggests that subjects could no longer use learned "responses" as a route to action and that neural priming in these regions during the priming phase reflects stimulus-response learning rather than priming of conceptual information. (Adapted with permission from I. G. Dobbins, D. M. Schnyer, M. Verfaellie, & D. L. Schacter, 2004, Cortical activity reductions during repetition priming can result from rapid response learning, *Nature*, *428*, 316–319. Copyright 2004, reprinted by permission from Macmillan Publishers, Ltd.)

The role of response learning in conceptual priming tasks has received support from a study by Dobbins, Schnyer, and colleagues (2004). In this study (figure 48.9), stimuli (e.g., "Bulldozer") were repeatedly semantically classified (e.g., "Larger than a shoebox?"), with the specific classification decision and the corresponding response either being held constant across repetitions or changed across repetitions (e.g., "Smaller than a shoebox?"). While repetition of a stimulus with the identical decision cue was associated with robust repetition suppression in left VLPFC, repetition of a stimulus with the inverted decision cue was associated with diminished repetition suppression in this region. Because the same conceptual information is accessed across the decision cues, the disruption of priming with cue inversion suggests that the left VLPFC repetition suppression effects typically observed in conceptual priming tasks are at least partially attributable to stimulus-response learning rather than priming of conceptual information. While these data provide an important challenge to accounts of left VLPFC priming that focus only on the reduction in cognitive control demands following cortical tuning of semantic representations, one caveat is that the design used by Dobbins and colleagues covaried repetition at the "decision" and "response" levels. That is, switching the decision from "Larger than a shoebox?" to "Smaller than a shoebox?" requires both a decision switch and a response switch (Schacter et al., 2004; Schnyer, Dobbins, Nicholls, Schacter, & Verfaellie, 2006). Indeed, behavioral evidence suggests that priming at the decision level can be dissociated from response repetition (Schnyer et al., 2007).

Together, extant evidence suggests that prior conceptual processing can reduce demands on PFC control mechanisms during future conceptual processing. However, additional work is needed to establish the extent to which these PFC activation reductions reflect priming at different levels of processing (i.e., conceptual, decision, or response). An intriguing hypothesis is that these distinct levels of learning might give rise to dissociable forms of neural priming. For example, priming at the conceptual level may reduce demands on processing in left anterior VLPFC—a region that has repeatedly been implicated in controlled semantic retrieval—whereas learning at the response level may reduce demands on processing in regions more directly related to response selection (e.g., premotor areas) (Race, Shanker, & Wagner, 2008). Of additional interest is whether these distinct forms of priming—from higher-level conceptual priming to lower-level response learning—correspond to a representational hierarchy within PFC (Fuster, 2001; Badre & D'Esposito, 2007; Koechlin & Summerfield, 2007), perhaps organized along an anterior (higher-level) to posterior (lower-level) gradient (Race et al., 2008). Insight into these questions will provide a more complete understanding of the multiple ways in which learning from past experiences can reduce future uncertainty, and thus demands on PFC-mediated control.

Conclusion

In this chapter we reviewed influential theories of cognitive control and considered the specific manner in which VLPFC control mechanisms serve to resolve interference and reduce uncertainty during mnemonic processing. While our focus on VLPFC operations reflects the considerable progress that has been made in understanding VLPFC function (for reviews, see Petrides, 2005; Badre & Wagner, 2007), it should be emphasized that other PFC mechanisms work in conjunction with VLPFC to achieve mnemonic goals (for reviews, see Fletcher & Henson, 2001; Wagner, 2002; Buckner, 2003; Simons & Spiers, 2003). For example, it has been argued that while VLPFC supports "active retrieval" of mnemonic representations, DLPFC subserves the complementary role of monitoring mnemonic representations once activated (Petrides, 1996, 2005). To the extent that DLPFC supports the monitoring of mnemonic information (Henson et al., 2000; Fletcher & Henson, 2001; Dobbins et al., 2002; Rugg et al., 2003; Achim & Lepage, 2005; Dobbins, Simons, et al., 2004), this argument would suggest a hierarchical, but interactive, relationship between VLPFC and DLPFC retrieval operations. Along similar lines, it has been suggested that VLPFC and DLPFC are hierarchically organized during episodic encoding, with VLPFC serving a general role in encoding (e.g., Wagner et al., 1998; Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998), but DLPFC selectively recruited when encoding involves processing the relationship between multiple stimuli (Blumenfeld & Ranganath, 2006; Murray & Ranganath, 2007). Further delineation of the contributions of DLPFC to mnemonic processing, as well as the nature of DLPFC-VLPFC interactions, remains an important avenue for future research. Finally, frontopolar cortex has frequently been implicated in higher-order forms of mnemonic processing (for reviews, see Rugg & Wilding, 2000; Fletcher & Henson, 2001; Buckner, 2003; Ramnani & Owen, 2004), though ambiguity remains concerning the specific nature of frontopolar interactions with "lower" forms of mnemonic control. Further advances in our understanding of the interplay between PFC control and mnemonic processing will require consideration of both the computations supported by specific PFC subregions and the manner in which coordinated processing across these subregions allows for the achievement of mnemonic goals.

ACKNOWLEDGMENTS This work was supported by grants from the National Institute of Mental Health (5R01-MH076932-02; 5R01-MH080309-02), the Alfred P. Sloan Foundation, and the National Alliance for Research on Schizophrenia and Depression.

REFERENCES

Achim, A. M., & Lepage, M. (2005). Dorsolateral prefrontal cortex involvement in memory post-retrieval monitoring revealed in both item and associative recognition tests. *NeuroImage*, *24*, 1113–1121.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford, UK: Oxford University Press.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *J. Mem. Lang.*, *49*, 415–445.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.*, *20*, 1063–1087.

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., Glover, G. H., & Gabrieli, J. D. (2004). Neural systems underlying the suppression of unwanted memories. *Science*, *303*, 232–235.

Aron, A. A., Durston, S., Eagle, D. M., Logan, G. D., Stinear, C. M., & Stuphorn, V. (2007). Converging evidence for a frontal-basal-ganglia network for inhibitory control of action and cognition. *J. Neurosci.*, *27*, 11860–11864.

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to Stop signal response inhibition: Role of the subthalamic nucleus. *J. Neurosci.*, *26*, 2424–2433.

Badgaiyan, R. D., Schacter, D. L., & Alpert, N. M. (2001). Priming within and across modalities: Exploring the nature of rCBF increases and decreases. *NeuroImage*, *13*, 272–282.

Badre, D. (2008). Cognitive control, hierarchy, and the rostrocaudal organization of the frontal lobes. *Trends Cogn. Sci.*, *12*, 193–200.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J. Cogn. Neurosci.*, *19*, 2082–2099.

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*, 907–918.

Badre, D., & Wagner, A. D. (2002). Semantic retrieval, mnemonic control, and prefrontal cortex. *Behav. Cogn. Neurosci. Rev.*, *1*, 206–218.

Badre, D., & Wagner, A. D. (2004). Selection, integration, and conflict monitoring: Assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron*, *41*, 473–487.

Badre, D., & Wagner, A. D. (2005). Frontal lobe mechanisms that resolve proactive interference. *Cereb. Cortex*, *15*, 2003–2012.

Badre, D., & Wagner, A. D. (2006). Computational and neurobiological mechanisms underlying cognitive flexibility. *Proc. Natl. Acad. Sci. USA*, *103*, 7186–7191.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*, 2883–2901.

Blumenfeld, R. S., & Ranganath, C. (2006). Dorsolateral prefrontal cortex promotes long-term memory formation through its role in working memory organization. *J. Neurosci.*, *26*, 916–925.

Botvinick, M. M. (2007). Multilevel structure in behaviour and in the brain: A model of Fuster's hierarchy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *362*, 1615–1626.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends Cogn. Sci.*, *8*, 539–546.

Braver, T. S., & Bongiolatti, S. R. (2002). The role of frontopolar cortex in subgoal processing during working memory. *NeuroImage*, *15*, 523–536.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, *39*, 713–726.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, *281*, 1185–1187.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*, 1118–1121.

Buckner, R. L. (2003). Functional-anatomic correlates of control processes in memory. *J. Neurosci.*, *23*, 3999–4004.

Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B., & Dale, A. M. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron*, *20*, 285–296.

Bunge, S. A., Burrows, B., & Wagner, A. D. (2004). Prefrontal and hippocampal contributions to visual associative recognition: Interactions between cognitive control and episodic retrieval. *Brain Cogn.*, *56*, 141–152.

Bunge, S. A., Ochsner, K. N., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (2001). Prefrontal regions involved in keeping information in and out of mind. *Brain*, *124*, 2074–2086.

Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2004). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cereb. Cortex*, *15*, 239–249.

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Curr. Dir. Psychol. Sci.*, *15*, 118–121.

Cabeza, R., Locantore, J. K., & Anderson, N. D. (2003). Lateralization of prefrontal activity during episodic memory retrieval: Evidence for the production-monitoring hypothesis. *J. Cogn. Neurosci.*, *15*, 249–259.

Carlesimo, G. A., Turriziani, P., Paulesu, E., Gorini, A., Caltagirone, C., Fazio, F., & Perani, D. (2003). Brain activity during intra- and cross-modal priming: New empirical data and review of the literature. *Neuropsychologia*, *42*, 14–24.

Christoff, K., & Gabrieli, J. D. E. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology*, *28*, 168–186.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, *14*, 1136–1149.

Christoff, K., Ream, J. M., Geddes, L. P., & Gabrieli, J. D. (2003). Evaluating self-generated information: Anterior prefrontal contributions to human cognition. *Behav. Neurosci.*, *117*, 1161–1168.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychol. Rev.*, *97*, 332–361.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychol. Rev.*, *99*, 45–77.

Danker, J. F., Gunn, P., & Anderson, J. R. (2008). A rational account of memory predicts left prefrontal activation during controlled retrieval. *Cereb. Cortex*, *18*, 2674–2685.

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *J. Neurosci.*, *15*, 5870–5878.

De Pisapia, N., Slomski, J. A., & Braver, T. S. (2007). Functional specializations in lateral prefrontal cortex associated with the integration and segregation of information in working memory. *Cereb. Cortex*, *17*, 993–1006.

Depue, B. E., Curran, T., & Banich, M. T. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. *Science*, *317*, 215–219.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. USA*, *93*, 13494–13499.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, *18*, 193–222.

D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., & Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Brain Res. Cogn. Brain Res.*, *7*, 1–13.

D'Esposito, M., Postle, B. R., Ballard, D., & Lease, J. (1999). Maintenance versus manipulation of information held in working memory: An event-related fMRI study. *Brain Cogn.*, *41*, 66–86.

D'Esposito, M., Postle, B. R., Jonides, J., & Smith, E. E. (1999). The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI. *Proc. Natl. Acad. Sci. USA*, *96*, 7514–7519.

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron*, *35*, 989–996.

Dobbins, I. G., Schnyer, D. M., Verfaellie, M., & Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from rapid response learning. *Nature*, *428*, 316–319.

Dobbins, I. G., Simons, J. S., & Schacter, D. L. (2004). fMRI evidence for separable and lateralized prefrontal memory monitoring processes. *J. Cogn. Neurosci.*, *16*, 908–920.

Dobbins, I. G., & Wagner, A. D. (2005). Domain-general and domain-sensitive prefrontal mechanisms for recollecting events and detecting novelty. *Cereb. Cortex*, *15*, 1768–1778.

Dolan, R. J., & Fletcher, P. C. (1997). Dissociating prefrontal and hippocampal function in episodic memory encoding. *Nature*, *388*, 582–585.

Fellows, L. K., & Farah, M. J. (2005). Is anterior cingulate cortex necessary for cognitive control? *Brain*, *128*, 788–796.

Feredoes, E., Tononi, G., & Postle, B. R. (2006). Direct evidence for a prefrontal contribution to the control of proactive interference in verbal working memory. *Proc. Natl. Acad. Sci. USA*, *103*, 19530–19534.

Fletcher, P. C., & Henson, R. N. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain*, *124*, 849–881.

Fletcher, P. C., Shallice, T., & Dolan, R. J. (2000). "Sculpting the response space"—An account of left prefrontal activation at encoding. *NeuroImage*, *12*, 404–417.

Fuster, J. M. (2001). The prefrontal cortex—An update: Time is of the essence. *Neuron*, *30*, 319–333.

Gabrieli, J. D. E., Desmond, J. E., Demb, J. B., Wagner, A. D., Stone, M. V., Vaidya, C. J., & Glover, G. H. (1996). Functional magnetic resonance imaging of semantic memory processes in the frontal lobes. *Psychol. Sci.*, *7*, 278–283.

Garavan, H., Ross, T. J., Li, S. J., & Stein, E. A. (2000). A parametric manipulation of central executive functioning. *Cereb. Cortex*, *10*, 585–592.

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: Functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *J. Neurosci.*, *26*, 6523–6532.

Gough, P. M., Nobre, A. C., & Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J. Neurosci.*, *25*, 8010–8016.

Green, A. E., Fugelsang, J. A., Kraemer, D. J., Shamosh, N. A., & Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Res.*, *1096*, 125–137.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cogn. Sci.*, *10*, 14–23.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *362*, 1601–1613.

Henson, R. N. (2003). Neuroimaging studies of priming. *Prog. Neurobiol.*, *70*, 53–81.

Henson, R. N. A., Rugg, M. D., Shallice, T., & Dolan, R. J. (2000). Confidence in recognition memory for words: Dissociating right prefrontal roles in episodic retrieval. *J. Cogn. Neurosci.*, *12*, 913–923.

Henson, R. N. A., Shallice, T., Josephs, O., & Dolan, R. J. (2002). Functional magnetic resonance imaging of proactive interference during spoken cued recall. *NeuroImage*, *17*, 543–558.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia*, *27*, 1043–1056.

Johansson, M., Aslan, A., Bäuml, K. H., Gabel, A., & Mecklinger, A. (2007). When remembering causes forgetting: Electrophysiological correlates of retrieval-induced forgetting. *Cereb. Cortex*, *17*, 1335–1341.

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, *139*, 181–193.

Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., & Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proc. Natl. Acad. Sci. USA*, *95*, 8410–8413.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, *399*, 148–151.

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science*, *318*, 594–598.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, *50*, 963–974.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*, 1181–1185.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.*, *11*, 229–235.

Kostopoulos, P., & Petrides, M. (2003). The mid-ventrolateral prefrontal cortex: Insights into its role in memory retrieval. *Eur. J. Neurosci.*, *17*, 1489–1497.

Kostopoulos, P., & Petrides, M. (2008). Left mid-ventrolateral prefrontal cortex: Underlying principles of function. *Eur. J. Neurosci.*, *27*, 1037–1049.

Kuhl, B. A., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nat. Neurosci.*, *10*, 908–914.

Kuhl, B. A., Kahn, I., Dudukovic, N. M., & Wagner, A. D. (2008). Overcoming suppression in order to remember: Contributions from anterior cingulate and ventrolateral prefrontal cortex. *Cogn. Affective Behav. Neurosci.*, *8*, 211–221.

Law, J. R., Flanery, M. A., Wirth, S., Yanike, M., Smith, A. C., Frank, L. M., Suzuki, W. A., Brown, E. N., & Stark, C. E. L. (2005). Functional magnetic resonance imaging activity during the gradual acquisition and expression of paired-associate memory. *J. Neurosci.*, *25*, 5720–5729.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends Cogn. Sci.*, *6*, 299–305.

Lundstrom, B. N., Ingvar, M., & Petersson, K. M. (2005). The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. *NeuroImage*, *27*, 824–834.

MacDonald, A. W., 3rd, Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–1838.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nat. Neurosci.*, *11*, 103–107.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.*, *24*, 167–202.

Murray, L. J., & Ranganath, C. (2007). The dorsolateral prefrontal cortex contributes to successful relational memory encoding. *J. Neurosci.*, *27*, 5515–5522.

Nee, D. E., Jonides, J., & Berman, M. G. (2007). Neural mechanisms of proactive interference-resolution. *NeuroImage*, *38*, 740–751.

Nolde, S. F., Johnson, M. K., & D'Esposito, M. (1998). Left prefrontal activation during episodic remembering: An event-related fMRI study. *NeuroReport*, *9*, 3509–3514.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.*, *18*, 283–328.

Owen, A. M., Evans, A. C., & Petrides, M. (1996). Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: A positron emission tomography study. *Cereb. Cortex*, *6*, 31–38.

Petrides, M. (1994). Frontal lobes and behaviour. *Curr. Opin. Neurobiol.*, *4*, 207–211.

Petrides, M. (1996). Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *351*, 1455–1461.

Petrides, M. (2000). The role of the mid-dorsolateral prefrontal cortex in working memory. *Exp. Brain Res.*, *133*, 44–54.

Petrides, M. (2002). The mid-ventrolateral prefrontal cortex and active mnemonic retrieval. *Neurobiol. Learn. Mem.*, *78*, 528–538.

Petrides, M. (2005). Lateral prefrontal cortex: Architectonic and functional organization. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *360*, 781–795.

Petrides, M. (2006). The rostro-caudal axis of cognitive control processing within lateral frontal cortex. In S. Dehaene, J.-R. Duhamel, M. D. Hauser, & G. Rizzolatti (Eds.), *From monkey brain to human brain: A Fyssen Foundation Symposium* (pp. 293–314). Cambridge, MA: MIT Press.

Petrides, M., & Pandya, D. N. (1999). Dorsolateral prefrontal cortex: Comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *Eur. J. Neurosci.*, *11*, 1011–1036.

Petrides, M., & Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur. J. Neurosci.*, *16*, 291–310.

Petrides, M., & Pandya, D. N. (2007). Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *J. Neurosci.*, *27*, 11573–11586.

Postle, B. R., Berger, J. S., & D'Esposito, M. (1999). Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proc. Natl. Acad. Sci. USA*, *96*, 12959–12964.

Race, E., Shanker, S., & Wagner, A. D. (2008). Neural priming in human frontal cortex: Multiple forms of learning reduce demands on the prefontal executive system. *J. Cogn. Neurosci.*, 1–16.

Raichle, M. A., Feiz, J. A., Videen, T. O., MacLeod, A. M. K., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human functional anatomy during non-motor learning. *Cereb. Cortex*, *4*, 8–26.

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.*, *5*, 184–194.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol. Sci.*, *17*, 249–255.

Roediger, H. L., III, & McDermott, K. B. (1993). Implicit memory in normal human subjects. In H. Spinnler & F. Boller (Series Eds.) & F. Boller & J. Grafman (Vol. Eds.), *Handbook of neuropsychology* (pp. 63–131). Amsterdam: Elsevier.

Rugg, M. D., Henson, R. N., & Robb, W. G. (2003). Neural correlates of retrieval processing in the prefrontal cortex during recognition and exclusion tasks. *Neuropsychologia*, *41*, 40–52.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends Cogn. Sci.*, *4*, 108–115.

Rypma, B., Prabhakaran, V., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Load-dependent roles of frontal brain regions in the maintenance of working memory. *NeuroImage*, *9*, 216–226.

Salmon, E., Van der Linden, M., Collette, F., Delfiore, G., Maquet, P., Degueldre, C., Luxen, A., & Franck, G. (1996). Regional brain activity during working memory tasks. *Brain*, *119*, 1617–1625.

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*, 185–195.

Schacter, D. L., Dobbins, I. G., & Schnyer, D. M. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nat. Rev. Neurosci.*, *5*, 853–862.

Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Curr. Opin. Neurobiol.*, *17*, 171–176.

Schnyer, D. M., Dobbins, I. G., Nicholls, L., Davis, S., Verfaellie, M., & Schacter, D. L. (2007). Item to decision mapping in rapid response learning. *Mem. Cogn.*, *35*, 1472–1482.

Schnyer, D. M., Dobbins, I. G., Nicholls, L., Schacter, D. L., & Verfaellie, M. (2006). Rapid response learning in amnesia: Delineating associative learning components in repetition priming. *Neuropsychologia*, *44*, 140–149.

Shimamura, A. P., Jurica, P. J., Mangels, J. A., Gershberg, F. B., & Knight, R. T. (1995). Susceptibility to memory interference effects following frontal lobe damage: Findings from tests of paired-associate learning. *J. Cogn. Neurosci.*, *7*, 144–152.

Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nat. Rev. Neurosci.*, *4*, 637–648.

Smith, M. L., Leonard, G., Crane, J., & Milner, B. (1995). The effects of frontal- or temporal-lobe lesions on susceptibility to interference in spatial memory. *Neuropsychologia*, *33*, 275–285.

Sohn, M. H., Goode, A., Stenger, V. A., Carter, C. S., & Anderson, J. R. (2003). Competition and representation during memory retrieval: Roles of the prefrontal cortex and the posterior parietal cortex. *Proc. Natl. Acad. Sci. USA*, *100*, 7412–7417.

Sohn, M. H., Goode, A., Stenger, V. A., Jung, K. J., Carter, C. S., & Anderson, J. R. (2005). An information-processing model of three cortical regions: Evidence in episodic memory retrieval. *NeuroImage*, *25*, 21–33.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc. Natl. Acad. Sci. USA.*, *94*, 14792–14797.

Thompson-Schill, S. L., D'Esposito, M., & Kan, I. P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron*, *23*, 513–522.

Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D'Esposito, M., Kan, I. P., Knight, R. T., & Swick, D. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cogn. Affective Behav. Neurosci.*, *2*, 109–120.

Thompson-Schill, S. L., Swick, D., Farah, M. J., D'Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *Proc. Natl. Acad. Sci. USA*, *95*, 15855–15860.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301–306.

Wagner, A. D. (2002). Cognitive control and episodic memory: Contributions from prefrontal cortex. In L. R. Squire & D. L. Schacter (Eds.), *Neuropsychology of memory* (3rd ed., pp. 174–192). New York: Guilford Press.

Wagner, A. D., Desmond, J. E., Demb, J. B., Glover, G. H., & Gabrieli, J. D. E. (1997). Semantic repetition priming for verbal and pictorial knowledge: A functional MRI study of left inferior prefrontal cortex. *J. Cogn. Neurosci.*, *9*, 714–726.

Wagner, A. D., & Koutstaal, W. (2002). Priming. In V. S. Ramachandran (Ed.), *Encyclopedia of the human brain* (vol. 4, pp. 27–46). San Diego: Academic Press.

Wagner, A. D., Maril, A., Bjork, R. A., & Schacter, D. L. (2001). Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal cortex. *NeuroImage*, *14*, 1337–1347.

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*, 329–338.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*, 1188–1191.

Walton, M. E., Devlin, J. T., & Rushworth, M. F. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nat. Neurosci.*, *7*, 1259–1265.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.*, *8*, 227–233.

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nat. Rev. Neurosci.*, *4*, 139–147.

49 Phases of Influence: How Emotion Modulates the Formation and Retrieval of Declarative Memories

ELIZABETH A. KENSINGER

ABSTRACT We tend to remember emotional experiences long after we have forgotten more mundane ones. The beneficial effects of emotion on memory appear to arise through influences between emotion-specific processes and domain-general sensory and mnemonic processes. These interactions arise at every phase of memory, including encoding, consolidation, and retrieval. As this chapter describes, emotion heightens perception and attention during encoding and enhances the likelihood that information is elaborated and organized. Emotion also modulates postencoding consolidation processes, increasing the likelihood that an emotional event is maintained in a durable memory trace. Emotion continues to wield its influence at retrieval, increasing the likelihood that information is retrieved and also augmenting the subjective vividness associated with the retrieved memory. This chapter discusses the neural processes that underlie these effects of emotion on memory. Particular emphasis is placed on understanding the role of the amygdala in emotional memory and how the amygdala exerts its effects by means of interactions with other sensory and mnemonic regions.

Events often elicit short-lived cognitive, physiological, and somatic reactions, otherwise known as emotions (see Barrett, 2006; Izard, 2007; Frijda & Sundararajan, 2007; Panksepp, 2007; Scherer, 2000, for discussion of the best way to think about the term). Emotional reactions accompany many of life's experiences, particularly those we care most about remembering. It is, therefore, critical to understand how emotion influences memory processes, as without this knowledge, it would be nearly impossible to discern how memory operates in everyday life. This realization has sparked interest in the study of "emotional memory," or the examination of how memories for experiences that triggered an emotional response are formed and retrieved.

Extensive research on emotional memory demonstrates that emotional experiences tend to be remembered better than experiences that lack emotional importance, an effect referred to as "emotional memory enhancement" (reviewed by Buchanan & Adolphs, 2004). This mnemonic benefit conveyed by emotion has long been acknowledged (see Colgrove, 1899, for a study examining memory for the assassination of President Abraham Lincoln), but it is only within recent decades that research has begun to elucidate the processes that give rise to it. Though animal research has clarified many of the mechanisms that support emotion's influence on memory (reviewed by McGaugh, 2004; Phelps & LeDoux, 2005), this chapter focuses exclusively on the effects of emotion on declarative memory in humans, describing how behavioral, neuropsychological, neuropharmacological, and neuroimaging studies have elucidated how emotion influences each phase of memory (see figure 49.1). Because the study of emotional memory in humans is still a relatively young topic of investigation, this chapter concludes with a discussion of directions for future research, including the need to consider individual differences when assessing the effects of emotion on memory.

ELIZABETH A. KENSINGER Department of Psychology, Boston College, Chestnut Hill; Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, Massachusetts

The influence of emotion during encoding

It is well known that the way in which information is processed initially has downstream consequences on the likelihood that the information is remembered later (Craik & Lockhart, 1972), with information that is detected, attended, and elaborated upon being the most likely to be remembered (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996). Many of emotion's effects on memory appear to arise through broader influences on the way in which emotional information is detected and attended at the outset. Emotional stimuli are noticed more quickly and more often than nonemotional ones (Anderson, 2005; Fox, Russo, Bowles, & Dutton, 2001; Leclerc & Kensinger, 2008; Ohman, Flykt, & Esteves, 2001; Phelps, Ling, & Carrasco, 2006; Williams, Mathews, & MacLeod, 1996), and the processing of emotional information is prioritized so that it can occur even when attentional resources are taxed (reviewed by Dolan & Vuilleumier, 2003; Pessoa, 2005; Vuilleumier & Driver, 2007). Once an emotional item is detected, attention

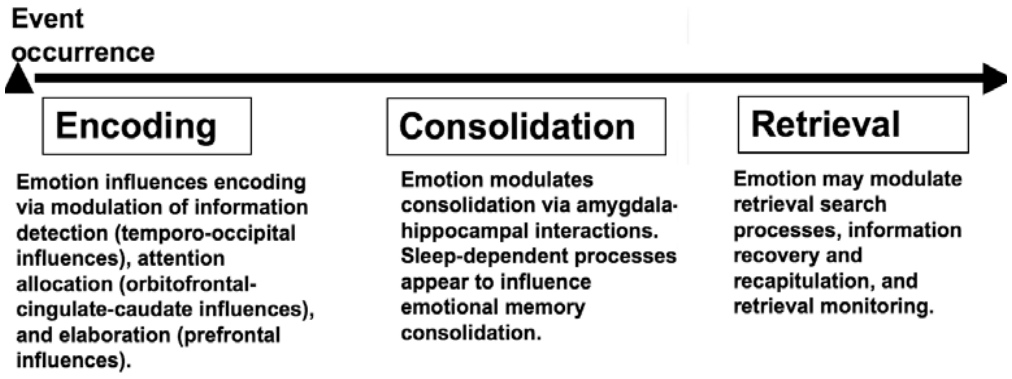


FIGURE 49.1 Overview of the effects of emotion on memory.

also is more likely to be focused and sustained on it (e.g., Armony & Dolan, 2002; Mogg, Bradley, de Bono, & Painter, 1997), and individuals are more likely to elaborate on the emotional information, connecting it with existing semantic or autobiographical information (e.g., Buchanan, Etzel, Adolphs, & Tranel, 2006; Talmi & Moscovitch, 2004; Talmi, Schimmack, Paterson, & Moscovitch, 2007). Each of these factors can increase the likelihood that emotional information is encoded into a stable memory trace. In fact, Talmi, Luk, McGarry, & Moscovitch (2007) have proposed that direct modulation of memory may not be required for short-term enhancements in the retention of emotional information. Rather, emotion's modulation of domain-general processes—enhanced attention, distinctive encoding, and information elaboration and organization—may be sufficient to mediate emotion's benefit on retention of information over relatively short delays. As will be described subsequently, neuroimaging may provide one means to clarify the extent to which emotion's influence on memory is mediated through influences on information processing rather than dependent on direct modulation of memory binding and consolidation processes (see also Talmi, Anderson, Riggs, Caplan, & Moscovitch, 2008).

THE EFFECT OF EMOTION ON INFORMATION DETECTION AND ATTENTION ALLOCATION Many of emotion's influences on detection and attention appear to arise through interactions between the amygdala and other sensory regions. It is proposed that once the amygdala is activated by emotional stimuli, it can modulate the functioning of sensory cortices to assure that emotional information is attended (LeDoux, 1995). This hypothesis is anatomically plausible, because the amygdala has strong reciprocal connections with most sensory regions (Amaral, Price, Pitkanen, & Carmichael, 1992; Amaral, 2003). The hypothesis also is supported by neuroimaging studies that reveal strong correlations between the amount of activity in the amygdala and in visual processing regions including the fusiform gyrus (e.g., Noesselt, Driver, Heinze, & Dolan, 2005; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004) and occipital lobe (Tabert et al., 2001; figure 49.2*A*) during the processing of emotional information. Although these correlations cannot establish the directionality of the modulation, they are consistent with the proposal that the amygdala can modulate sensory functioning. Stronger evidence for an amygdala-mediated influence on sensory activity came from a study in which Vuilleumier and colleagues (2004) asked individuals with varying amounts of amygdala damage to view fearful and neutral faces while in an fMRI scanner. Only patients with a functioning amygdala showed fusiform modulation in response to the facial expression, with greater fusiform activity to fearful than to neutral faces. In fact, there was a strong correlation between the amount of intact amygdala and the amount of fusiform modulation in response to the fearful faces, consistent with the proposal that the amygdala has a modulatory effect on visual processing regions, increasing the likelihood that emotional information is detected and processed.

Interactions between the amygdala and sensory regions also seem to enhance memory for the visual details of emotional stimuli (Mickley & Kensinger, 2008; Kensinger, Garoff-Eaton, & Schacter, 2007b). Participants are more likely to remember the precise visual attributes of a negative item as compared to a neutral one; for example, they recognize exactly which grenade they have seen more often than they recognize which blender they have seen (Kensinger, Garoff-Eaton, & Schacter, 2006). This effect appears to arise from interactions between the amygdala and the fusiform gyrus during encoding. As compared to the processing of neutral items, during the processing of negative items that will later be remembered with precise visual detail, there is increased activity in the amygdala and the right fusiform gyrus. There also is a strong correlation between the amount of activity in these two regions during the processing of negative items, whereas no such correlation exists during the

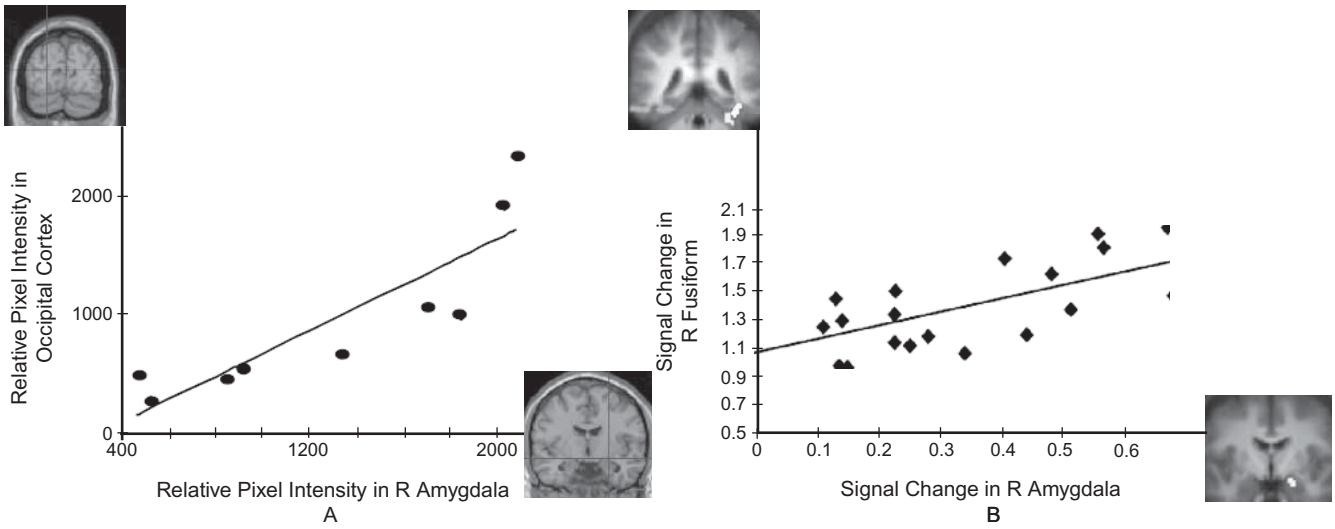


FIGURE 49.2 During the processing of negatively emotional information, there often are robust correlations between amygdala activity and activity in sensory processing regions (panel *A*, adapted from Tabert et al., 2001; images depict coordinates reported in that paper). These correlations are particularly strong during the encoding of negative items that will later be remembered with specific visual details (panel *B*, data from Kensinger, Garoff-Eaton, & Schacter, 2007b).

processing of neutral items (Kensinger et al., 2007b; figure 49.2*B*). The right fusiform gyrus is a region that is associated with the processing of visually specific details (e.g., Koutstaal et al., 2001) and with subsequent memory for the visual details of neutral items (Garoff, Slotnick, & Schacter, 2005). Therefore, it makes sense that enhanced activity within this region could increase the likelihood that the visual details of a negatively emotional item are remembered.

Modulation of sensory processes does not appear to be the only avenue by which emotion enhances memory for visual detail. The ability to remember the visual details of emotional items also may be tied to the way in which attention is allocated during encoding. Emotion does not seem to uniformly enhance memory for all aspects of an experience. Rather, some event details are remembered well and others are readily forgotten (reviewed by Buchanan & Adolphs, 2002; Kensinger, 2007; Mather, 2007; Reisberg & Heuer, 2004). For example, when presented with complex visual scenes, it often is that case that the visual details of the emotional aspects are remembered well but the visual details of the nonemotional aspects are remembered poorly (e.g., Kensinger, Garoff-Eaton, & Schacter, 2007a; Payne, Stickgold, Swanberg, & Kensinger, 2008; figure 49.3). An fMRI study revealed that activity in an affective-attentional network, including the right orbitofrontal cortex, the anterior cingulate gyrus, and the caudate nucleus, corresponds with the ability to remember the visual details of an emotional item but also with the *in*ability to remember other aspects associated with the item's presentation, such as what decision a person made about an item (Kensinger et al., 2007b; figure 49.4). This finding suggests that emotional items may be remembered in a detailed fashion because attention is focused on the intrinsic details of those items; however, by focusing on those emotional elements, other event details may be missed or easily forgotten.

It is interesting to note that this attentional focusing on emotional items does not arise through engagement of the same frontoparietal attention circuits that guide attention toward nonemotional, task-relevant information (reviewed by Corbetta & Shulman, 2002). Rather, when attention is focused on emotional information, it appears to be through engagement of emotion-specific processes that are brought online when a task requires engagement of motivational processes and of attention to affective stimuli (e.g., Robbins & Everitt, 1996; Schultz, 2000). This dissociation suggests that emotional information may be attended as a result of the engagement of emotion-specific processes rather than the domain-general ones that guide attention toward any task-relevant piece of information (and see Vuilleumier & Driver, 2007, for further discussion). Thus, when an individual is affectively focused on an item, attention appears to be drawn to the intrinsic attributes of that emotional item. This selective attention seems to have downstream mnemonic consequences, leading those intrinsic item details to be remembered better than elements only extrinsically linked to the emotional item (see Kensinger, 2007; Mather, 2007, for further discussion). These findings highlight that even when emotion's effects on memory seem to be mediated by influences on domain-general processes (such as attention allocation), this mediation may actually reflect emotion-specific modulation of sensory and attentional processes.

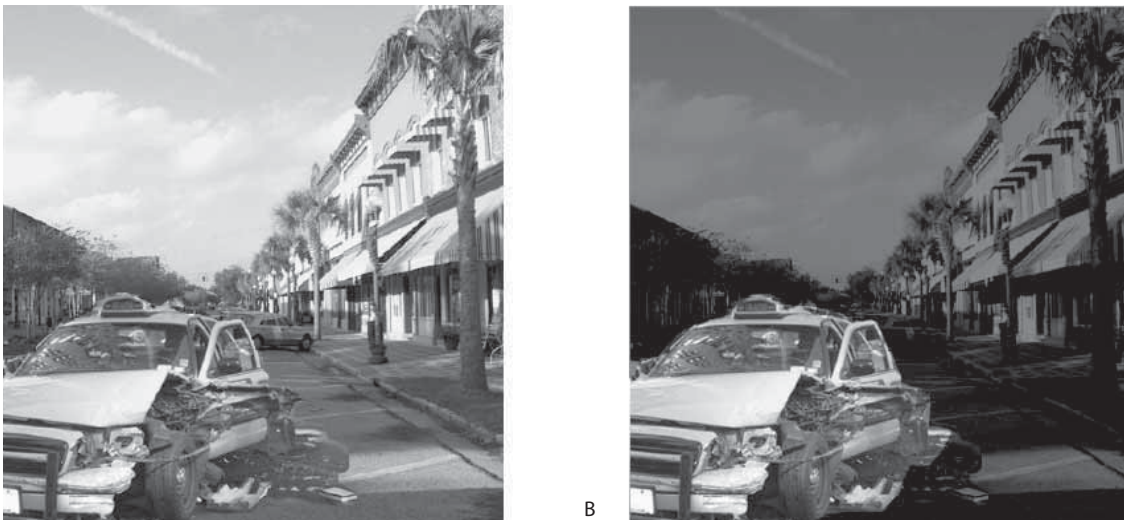


FIGURE 49.3 After experiencing an emotional event—such as a car accident (*A*)—participants may retain good memory for the details of the accident itself, but poor memory for the contextual details, such as what the street looked like (*B*). Modulation of attentional focusing at encoding, as well as of consolidation processes, appears to contribute to this effect.

EFFECTS ON ELABORATION AND ORGANIZATION OF INPUT

In addition to these effects of emotion on information detection and attention allocation, emotion also appears to influence the likelihood that information is elaborated and organized. It is well known that events that elicit negative emotions are elaborated and rehearsed more often than events that elicit no emotion (reviewed by Ochsner & Schacter, 2003). Emotional items also benefit from organizational clustering to a greater degree than do nonemotional items (Buchanan, Etzel, et al., 2006; Talmi & Moscovitch, 2004; Talmi, Schimmack, et al., 2007). Because information that is elaborated and well organized is more likely to be remembered (Craik & Lockhart, 1972), it makes sense that if emotion provides an organizing structure or a basis for elaboration, these features would convey benefits to memory. Indeed, a number of behavioral studies have demonstrated an important role for elaborative and organizational processes in enhancing emotional memory (Phelps et al., 1998; Talmi & Moscovitch, 2004), and neuroimaging studies have confirmed that regions implicated in elaborative processing—including the lateral prefrontal cortex—often are disproportionately recruited during the successful encoding of emotional information (e.g., Dolcos, LaBar, & Cabeza, 2004; Kensinger & Corkin, 2004; Maratos, Allan, & Rugg, 2000). These elaborative processes appear to be particularly essential for boosting the encoding of emotional information that is not highly arousing (e.g., Buchanan, Etzel, et al., 2006; Bush & Geer, 2001; Kensinger & Corkin, 2003; Talmi, Schimmack, et al., 2007), perhaps because these items do not benefit from the same amygdala-mediated enhancements in detection and attention (reviewed by Kensinger, 2004). Thus, when elaborative processes cannot be engaged easily (for example, when attention is divided), the mnemonic enhancement for nonarousing emotional items disappears (Bush & Geer, 2001; Kensinger & Corkin, 2004; Kern, Libkuman, Otoni, & Holmes, 2005).

CONCLUSIONS ABOUT ENCODING

Nearly all studies examining memory for emotional information have revealed a strong correlation between how active the amygdala is during the encoding of emotional information and how well that emotional information is remembered (reviewed by Hamann, 2001; LaBar & Cabeza, 2006). These correlations exist both across participants (e.g., Cahill et al., 1996) and within a single participant (e.g., Canli, Zhao, Brewer, Gabrieli, & Cahill, 2000). They arise in tasks using verbal stimuli (e.g., Erk et al., 2003; Kensinger & Corkin, 2004), slide shows (e.g., Cahill et al., 1996), facial expressions (e.g., Sergerie, Lepage, & Armony, 2006), and colored photographs (e.g., Dolcos et al., 2004; Sharot, Delgado, & Phelps, 2004), and they hold across a range of encoding tasks. Though many of these studies have focused on the amygdala's modulation of hippocampal binding and consolidation processes (an issue we will return to in the next section), the studies reviewed have revealed that emotion can exert many of its influences on memory by means of alterations in earlier stages of information processing. In particular, engagement of emotion-specific processes, implemented by the amygdala and orbitofrontal cortex, can influence memory by modulating sensory and conceptual processes. These interactions can ensure that emotional information is detected, attended, and elaborated (see also Duncan & Barrett, 2007; Talmi et al., 2008). Thus, regardless of the particular stimuli, encoding instructions, or task design, the engagement of emotion-specific processes during

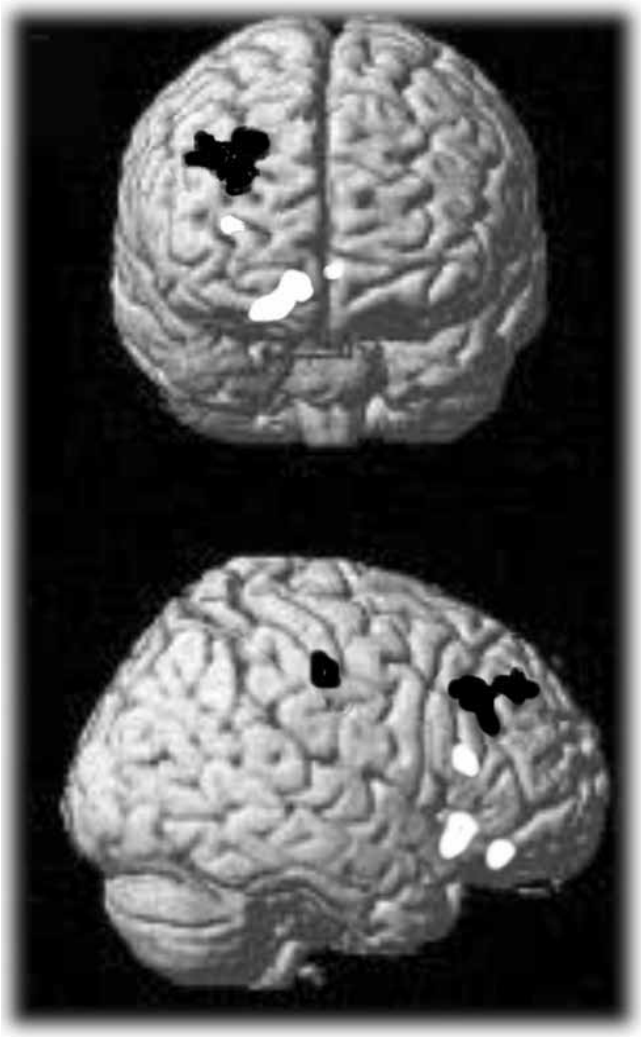


Figure 49.4 When participants view items and are asked to make a decision about them (e.g., to decide either whether it is alive or whether it is commonly encountered), participants often have good memory for the visual details of the item but poor memory for which decision they made about the item. Encoding-related activity can predict which items will later be remembered with visual detail but without memory for the decision. The activity that predicts this pattern of memory performance differs based on whether the items are negative (white regions) or neutral (black regions). This dissociation suggests that the narrowing of attention on intrinsic item attributes (such as an item's visual details) is mediated by different processes when the items are emotional versus nonemotional. (Adapted from Kensinger, Garoff-Eaton, & Schacter, 2007b.)

encoding is critically tied to the emotional memory enhancement effect.

Emotion influences consolidation processes

The effects of emotion on encoding processes can explain why emotional information is better remembered than neutral information across a range of delays. If emotional information is more likely to be detected, attended, and elaborated, then it would make sense that the information would be remembered well after both short and long delays. Interestingly, however, the effects of emotion often become exaggerated after a long delay (e.g., Kleinsmith & Kaplan, 1963; Walker & Tarte,1963; Sharot, Verfaellie, & Yonelinas, 2007; Sharot & Yonelinas, 2008), and damage to the amygdala tends to disproportionately influence the retention of emotional information over long delays while having a lesser influence on the ability to retain emotional information for only a short duration of time (Phelps, LaBar, & Spencer, 1997; Phelps et al., 1998; LaBar & Phelps, 1998). These findings cannot easily be explained by effects of emotion on encoding processes. Rather, these results point to the ability of emotion to influence the likelihood that memories are solidified into stable long-term traces. If emotion—through the actions of the amygdala—serves to increase the probability that a memory is consolidated, then it should follow that the benefit for emotional compared to nonemotional memories increases as the retention interval lengthens and that amygdala damage disrupts this effect. Indeed, as will be described later, there is abundant evidence to suggest that the amygdala modulates consolidation processes, enhancing the likelihood that an emotional memory can be remembered after a long delay.

Amygdala Modulation of Hippocampal Consolidation Extensive evidence for emotion's effects on hippocampal consolidation comes from animal studies of emotional learning. These studies reveal that stress hormones released as part of an affective response can trigger a cascade of neurotransmission and hormone release, ultimately leading to amygdalar modulation of hippocampal consolidation (reviewed by McGaugh, 2004; Phelps, 2004; Phelps & LeDoux, 2005). Paralleling these animal findings, research in humans suggests that emotion conveys benefits on the hippocampal consolidation of declarative memories. In particular, neuroimaging studies provide strong evidence for interactions between the amygdala and the hippocampus. The amygdala and hippocampus often are coactivated during the successful encoding of emotional information (e.g., Dolcos et al., 2004; Kensinger & Corkin, 2004), and there is a strong correlation between the activity in these regions (Kensinger & Corkin, 2004; Kensinger & Schacter, 2005a; Richardson, Strange, & Dolan, 2004), as well as an increase in functional connectivity between the regions (Kilpatrick & Cahill, 2003), as individuals learn emotional information. Although these studies cannot determine whether such interactions signify influences on memory consolidation per se, or on other mnemonic processes implemented by the hippocampal formation, the results are consistent with a modulatory role on consolidation processes.

The importance of consolidation processes in emotional memory also is suggested by the fact that many of the interactions between the amygdala and the hippocampus seem to be mediated by the influence of stress hormones. Administration of cortisol enhances the long-term recall of emotional information (Buchanan & Lovallo, 2001), whereas administration of beta blockers eliminates the enhancement (Cahill, Babinsky, Markowitsch, & McGaugh, 1995). Moreover, amygdala activation without an accompanying arousal response does not appear sufficient for a hippocampal-mediated boost in memory (e.g., Anderson, Yamaguchi, Grabski, & Lacka, 2006; Kensinger & Corkin, 2004), suggesting that both amygdala activation and a concurrent stress response are required. Although these studies cannot conclusively demonstrate an influence upon consolidation processes, because the amygdala's influence on hippocampal consolidation processes is thought to be modulated by release of stress hormones (McGaugh, 2004), these neuropharmacological and neuroimaging findings are consistent with a role for arousal-dependent amygdalar modulation of consolidation.

Emotion, Sleep, and Memory Consolidation Some of the mnemonic effects of emotion seem to be exerted through sleep-dependent consolidation processes. Although the details of how sleep influences memory consolidation are still debated (see Born, Rasch, & Gais, 2006; Ellenbogen, Hulbert, Stickgold, Dinges, & Thompson-Schill, 2006; Frank & Benington, 2006), extensive evidence suggests that sleep can benefit performance on both implicit and explicit memory tasks (reviewed by Payne, Ellenbogen, Walker, & Stickgold, in press; Stickgold, 2005; Walker & Stickgold, 2006). Sleep after learning can increase the rate of information acquisition (Born et al.; Gais, Lucas, & Born, 2006) and can make information less prone to interference and decay over time (Ellenbogen, Payne, & Stickgold, 2006). By contrast, sleep deprivation can impair declarative learning, suggesting that sleep is necessary for optimal hippocampal-dependent memory consolidation (C. Smith & Rose,1996). Slow-wave sleep may be particularly essential to the consolidation of episodic memories, with the reactivation and redistribution of hippocampal neural ensembles occurring during this sleep phase (reviewed by Marshall & Born, 2007; Rasch & Born, 2008).

Although only a few studies have compared the effects of sleep on memory for emotional versus nonemotional information, all of these studies supply evidence that sleep provides particular benefits for emotional memory. In one such study, Wagner, Hallschmid, Rasch, and Born (2006) asked participants to study narratives that were either emotional or neutral in content. Some participants slept for three hours after reading the narratives while other participants remained awake. Four years later, the benefits of sleep on emotional memory were apparent: sleep had no influence on the likelihood of remembering the topics of the neutral narratives, but the participants who slept after reading the narratives were more likely to remember the topics of the emotional narratives than were those who remained awake. Thus sleep conferred a particular mnemonic benefit for the emotional information.

We need not wait multiple years to see the benefits that sleep conveys to emotional memory. A few studies have revealed that emotional memory is better either after a brief (3-hour) period of REM-intensive sleep or after a full night of sleep rather than after a similar period of wakefulness (e.g., Hu, Stylos-Allan, & Walker, 2006; Wagner, Gais, & Born, 2001; Wagner et al., 2006; Wagner, Kashyap, Diekelmann, & Born, 2007). For example, Hu and colleagues presented participants with colored photographs from the International Affective Picture System (Lang, Bradley, & Cuthbert, 1999). Some photographs were emotionally arousing, and others were neutral. Participants' memory for the photos was tested 12 hours later, and the critical finding was that while sleep had no impact on memory for the neutral pictures, memory for the arousing pictures was better after a night of sleep than after an equivalent period of time spent awake.

Although sleep conveys benefits to memory for emotional items, it does not appear to enhance memory for all aspects of those items equally well. Rather, sleep's memory-enhancing effects appear to be specific to the most emotional aspects of a stimulus. When Payne, Stickgold, and colleagues (2008) presented participants with scenes including an emotional object placed on a neutral background (e.g., a snake in a forest), they found that sleep selectively preserved memory for the emotional object while conveying no memory benefit for the accompanying background (figure 49.5). This finding is intriguing, suggesting that not all aspects of an emotional event are consolidated as a single, bound entity. Rather, sleep appears to preferentially preserve memory for those elements of an experience that are strongly tied to the emotional nature of the event. This result suggests that there is an important interaction between emotion and sleep-mediated consolidation processes, with sleep having the greatest preservative benefit on memory for emotional information.

Concluding Remarks Regarding Consolidation There is no question that emotion modulates the consolidation of memories, with emotional experiences being more likely to be retained over time than nonemotional ones. What is less clear, however, is what attributes of an emotional experience benefit from enhanced consolidation. Within the emotional memory literature, there is an increasing

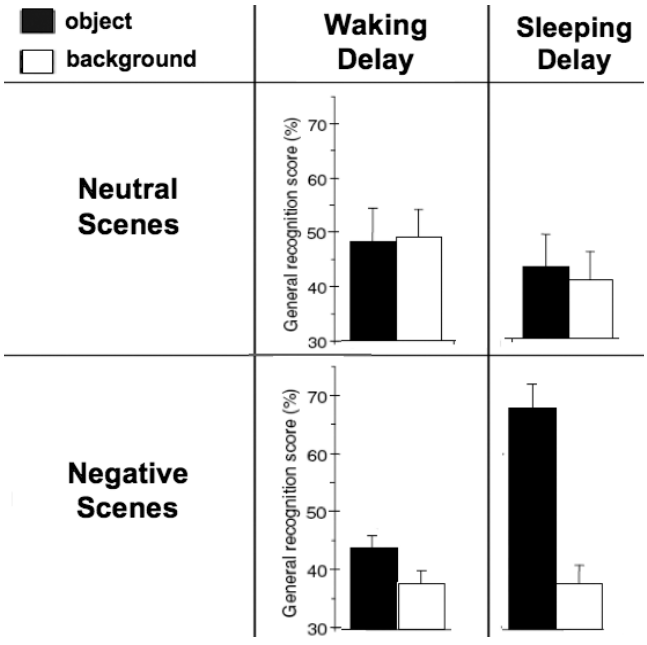


FIGURE 49.5 When participants study visual scenes and are tested on their memory for those scenes after either a 12-hour delay including a night of sleep or a 12-hour period of time spent awake, memory for the negative objects within scenes is selectively enhanced across a sleeping as compared to a waking delay. Memory for the backgrounds of those same scenes is unaffected by whether the delay included time spent awake or time spent asleep. (Data from Payne, Stickgold, Swanberg, & Kensinger, 2008.)

appreciation for the fact that not all aspects of an emotional experience are equally likely to be remembered (e.g., Reisberg & Heuer, 2004; Kensinger, 2007; Mather, 2007). Although many of these selective effects likely arise from attentional focusing at encoding (as discussed in the previous section), some of the effects may also arise through focal influences on consolidation. As noted earlier, sleep does not appear to benefit consolidation of all event attributes to an equal degree; rather, the benefits seem to be particularly pronounced for those details that are intrinsic to the emotional items. Further research is needed to reveal the extent to which encoding processes versus postencoding consolidation mechanisms lead to the focal enhancements in emotional memory, leading only some attributes of an emotional event to be remembered well.

The influence of emotion during retrieval

In comparison to the extensive number of studies that have examined the effects of emotion on encoding and consolidation processes, relatively few studies have investigated the influence of emotion during memory retrieval. However, the extant data indicate that emotion can modulate retrieval processes. The role of emotion in memory retrieval recently has been reviewed thoroughly (by Buchanan, 2007), and so here I will focus on the specific question of the amygdala's role during retrieval.

It is well established that the amygdala is active during the retrieval of emotional memories. Neuroimaging studies have revealed that amygdala engagement occurs both when the retrieval cue itself is emotional (e.g., Dolan, Lane, Chua, & Fletcher, 2000; Kensinger & Schacter, 2005b) and when the cue is neutral but the associated study context is emotional (e.g., Maratos, Dolan, Morris, Henson, & Rugg, 2001; Smith, Henson, Dolan, & Rugg, 2004; A. Smith, Henson, Rugg, & Dolan, 2005; Somerville, Wig, Whalen, & Kelley, 2006; Sterpenich et al., 2006). In one study, participants were asked to view objects that were presented against either neutral or emotional backgrounds. During recognition, they were shown the objects in isolation, and they had to indicate whether each object had been studied previously. The critical finding was that amygdala activity was greater during retrieval of items that had been studied with an emotional context than during retrieval of items that had been studied with a nonemotional context (A. Smith et al., 2004). The fact that amygdala activity was influenced by the study context, even when the retrieval cue itself was neutral, suggests that amygdala engagement during retrieval may not merely represent an emotional response to a retrieval cue. Rather, amygdala activity may be directly tied to the recovery of emotionally relevant information present during the encoding episode.

Though neuroimaging studies indicate that the amygdala is involved in the retrieval of emotional memories, they cannot speak to the necessity of the region. Indeed, there have been extensive discussions about whether the amygdala is essential for the retrieval of emotional memories (see Nader, 2003; LeDoux, 2000, for discussion). At least with regard to the retrieval of emotional autobiographical memories, recent patient studies have provided evidence that this region does play an essential role. Patients with damage to the amygdala have difficulty retrieving emotional memories, even of events that were experienced prior to the onset of their amygdala damage (Buchanan, Tranel, & Adolphs, 2005, 2006). Even when they do recall emotional experiences, patients with amygdala lesions rate them as being less emotional, as well as less vivid, than do control participants, suggesting that without the amygdala, emotional memories cannot be remembered as often or with the same qualitative richness as with an intact amygdala.

These studies cannot clarify the specific role played by the amygdala during the retrieval of emotional memories. In particular, it is unclear whether the amygdala's retrieval-related activity leads to or is caused by successful retrieval. It is widely accepted that memory retrieval consists of at

least a few distinct processes. After we receive a cue—passing someone familiar in the hallway, for example—we implement search processes to help us efficiently narrow down and sift through the information we have stored in mind in order to generate the sought-after information—such as where we last saw the person. If the search process is successful, the desired information will be recovered. After information is recovered, we assess whether it is what we were searching for, by engaging in retrieval monitoring processes. We may evaluate the plausibility of the retrieved information ("Could we really have seen this colleague on our flight back from Australia?") or consider our confidence in our memory ("Are we certain enough to mention the turbulent flight?"). Superimposed on all these processes, our brains seem to configure themselves into a "retrieval mode," allowing us to optimally query and evaluate the contents of our memory in order to reexperience past events (discussed by Rugg & Wilding, 2000; Sakai, 2003). Emotion could influence any of these stages of retrieval, and a great deal of ongoing research is investigating at which of these many phases emotion intervenes.

The amygdala may guide the search processes that lead to successful recovery of information. As noted earlier, patients with amygdala damage seem to select emotional memories less often than control participants (Buchanan, Tranel, & Adolphs, 2005, 2006). This finding is consistent with the hypothesis that amygdala damage results in altered mnemonic search process. In particular, if the amygdala typically boosts the efficiency or efficacy of the search process, making it more likely that relevant emotional information is recovered, then without this amygdala-facilitated search process, it would make sense that these patients would not receive any benefit when searching for information tied to an emotional experience. Neuroimaging evidence also indicates that amygdala activity can occur early in the retrieval process, before a memory has been fully elaborated, suggesting a role in the search process (Daselaar et al., 2008). Thus these findings lend credence to the proposal that amygdala activity modulates the search processes that lead to the successful recovery of emotional memories.

In addition to a role in the retrieval search process, amygdala engagement during retrieval also may be a reflection of the successful recovery of information. Retrieval-related activity can be tied to recapitulation, or the reinstantiation of processes engaged during encoding. For example, when retrieving a word that was paired with a sound at encoding, activity in auditory cortex often is high, whereas when retrieving information associated with pictorial information at encoding, visual activity can be enhanced (Kahn, Davachi, & Wagner, 2004; Wheeler, Petersen, & Buckner, 2000; Nyberg, Habib, McIntosh, & Tulving, 2000; Vaidya, Zhao, Desmond, & Gabrieli, 2002). In the same way as this activity is presumed to reflect the recapitulation of sensory processes engaged during encoding, so might amygdala engagement during retrieval reflect the reinstantiation of emotional processes engaged during encoding. Support for this hypothesis comes from studies demonstrating that amygdala activity during retrieval can be higher when participants are asked to determine whether an item was encoded in an emotional context than when they are asked to focus on nonemotional aspects of the item's presentation (A. Smith et al., 2004, 2005; A. Smith, Stephan, Rugg, & Dolan, 2006). This finding may suggest that the amygdala activity reflects the reinstantiation of the emotional context in which an event was learned. The fact that similar limbic regions often are associated both with successful encoding and with accurate retrieval of information (e.g., Fenker, Schott, Richardson-Klavehn, Heinze, & Duzel, 2005; Kensinger & Schacter, 2005a, 2005b) also may suggest that these regions' retrieval-related activity reflects the bringing online of the emotional information present during encoding or results from the reexperiencing of the emotion elicited during encoding.

Though these studies suggest a role for the amygdala in search and recovery processes, it also is possible that amygdala engagement is tied to memory monitoring processes and to metamemory assessments. In ERP studies, emotion can modulate late-onset positive potentials, believed to correspond with postretrieval monitoring processes (A. Smith et al., 2004). Though these studies do not implicate the amygdala specifically, they certainly suggest that emotion is likely to influence monitoring processes. However, the specific influence of emotion on mnemonic monitoring has been debated. Some hypothesize that amygdala engagement may bias monitoring processes in such a way as to lead people to believe that they have retrieved a particularly vivid and detailed memory. By this account, amygdala activity at retrieval may inflate a person's confidence in a memory, leading to a disconnect between the subjective vividness of a memory and the objective amount of detail included in that memory (see Sharot et al., 2004). Although there is evidence that amygdala engagement at retrieval is associated with the subjective vividness of a memory (Dolcos, LaBar, & Cabeza, 2005; Sharot et al.), its activity also can be elicited specifically during accurate retrieval (Kensinger & Schacter, 2005b, 2007). For this reason, it does not appear that amygdala engagement at retrieval serves only to inflate a person's confidence in a memory. Nevertheless, it is possible that amygdala engagement modulates processes tied to retrieval monitoring as well as those tied to retrieval success. More generally, it seems likely that amygdala activity is both a cause and consequence of emotional memory retrieval; however, further research is needed to examine whether there are situations in which amygdala activity is more strongly tied to one aspect of retrieval than to another.

Conclusions About Retrieval Though it is clear that amygdala engagement enhances encoding processes and facilitates consolidation, it is more widely debated whether the amygdala confers a benefit upon emotional memory retrieval. There is some evidence to suggest that limbic engagement primarily inflates a person's confidence in a memory (e.g., Sharot et al., 2004); but there is other evidence that limbic engagement at retrieval may be tied to remembering event details (e.g., Kensinger & Schacter, 2005b, 2008; A. Smith et al., 2006). It seems likely that, just as with its modulation of encoding and consolidation processes, the amygdala's influence during retrieval may critically depend on the types of details that a person is trying to recover. Perhaps amygdala engagement during retrieval facilitates the recovery of details intrinsically linked to an experience (e.g., the details of the emotional aspect of the event) but does not help with the recovery of details more peripheral to the elicited emotion (e.g., the nonemotional context in which the event occurred). Indeed, a study by Sharot, Martorella, Delgado, and Phelps (2007) revealed that enhanced amygdala activity during retrieval was associated with a reduction of activity in regions associated with retrieval of broader spatiotemporal context. It may be that when the amygdala is engaged, details intrinsic to the emotional aspects of the event are remembered, whereas retrieval of more peripheral, contextual details is impeded. Future research will do well to examine the validity of this hypothesis.

Concluding remarks and future directions

Emotion appears to influence the processes engaged during every phase of memory, but there are still many unanswered questions regarding how emotion exerts its influence. First, as alluded to in the preceding sections, we do not yet have a firm understanding of when emotion enhances, hinders, or exerts no influence on the likelihood of remembering information. It is well known that emotion does not lead to a picture-perfect memory (reviewed by Mather, 2007; Reisberg & Heuer, 2004). Nevertheless, emotional information—and particularly negative information—can be remembered with greater accuracy than nonemotional information (reviewed by Kensinger, 2007). Additional research is needed to understand which types of details are remembered well for emotional experiences and at which memory phases emotion conveys its mnemonic advantage. Future research will do well to investigate these issues not only through presentation of controlled stimuli within a laboratory setting, but also through assessment of participants' memories for emotional, autobiographical experiences.

Second, though researchers often assume that memory processes are consistent from one individual to the next, when it comes to emotion-memory interactions, there appear to be important individual differences. The sex of an individual can influence the neural processes that correspond with emotional memory enhancement, with men often showing more right-lateralized amygdala activity and women showing more left-lateralized amygdala activity (reviewed by Cahill, 2003; Hamann, 2005). Sex also can influence the magnitude of memory enhancement or memory trade-off elicited by emotion (discussed in Hamann, 2005). Personality characteristics, such as how neurotic someone is, also seem to influence the amount of amygdala activity elicited by stimuli (Hamann & Canli, 2004) and the likelihood that emotional information is detected (discussed by Duncan & Barrett, 2007), perhaps having downstream effects on the magnitude of emotional memory enhancement demonstrated. A person's level of anxiety or cognitive abilities also can influence emotional memory enhancement and the extent of mnemonic trade-off elicited when an emotional item is embedded in a nonemotional context (Waring, Payne, Schacter, & Kensinger, in press). A person's age also has fundamental influences on how emotional information is processed and remembered (reviewed by Kensinger & Leclerc, in press; Mather, 2006). These studies emphasize that research must examine not only how emotion impacts memory across all individuals, but also how individual differences influence the nature of emotion-memory interactions.

Third, as the resolution of MRI scans increases, it will be important for future research to move beyond thinking about the amygdala and the hippocampal memory system as single entities and to more thoroughly investigate how reciprocal influences are likely to depend on the particular subdivisions of each of these regions. Animal research has suggested that not all regions of the amygdala play the same modulatory role and that amygdalar interactions may not be equivalently strong with all medial temporal lobe structures (Davachi, 2006; McDonald, 2003). A finer appreciation of these anatomical distinctions within the human brain may go a long way toward revealing how emotion exerts its complex influences on memory formation, consolidation, and retrieval.

acknowledgments I thank Keely Muscatell, Jessica Payne, and Daniel Schacter for helpful discussion and for assistance in the preparation of this chapter. I gratefully acknowledge funding from the National Science Foundation (grant BCS-0542694) and the National Institute of Mental Health (grant MH080833).

REFERENCES

Amaral, D. G. (2003). The amygdala, social behavior, and danger detection. *Ann. NY Acad. Sci., 1000*, 337–347.

Amaral, D., Price, J., Pitkanen, A., & Carmichael, S. (1992). The amygdala: Neurobiological aspects of emotion, memory,

and mental dysfunction. In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (pp. 1–66). New York: Wiley-Liss.

Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *J. Exp. Psychol. Gen.*, *134*, 258–281.

Anderson, A. K., Yamaguchi, Y., Grabski, W., & Lacka, D. (2006). Emotional memories are not all created equal: Evidence for selective memory enhancement. *Learn. Memory*, *13*, 711–718.

Armony, J. L., & Dolan, R. J. (2002). Modulation of spatial attention by fear-conditioned stimuli: An event-related fMRI study. *Neuropsychologia*, *40*, 817–826.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspect. Psychol. Sci.*, *1*, 28–58.

Born, J., Rasch, B., & Gais, S. (2006). Sleep to remember. *Neuroscientist*, *12*, 410–424.

Buchanan, T. W. (2007). Retrieval of emotional memories. *Psychol. Bull.*, *133*, 761–779.

Buchanan, T. W., & Adolphs, R. (2002). The role of the human amygdala in emotional modulation of long-term declarative memory. In S. Moore & M. Oaksford (Eds.), *Emotional cognition: From brain to behavior* (p. 9–34). Amsterdam: John Benjamins.

Buchanan, T. W., & Adolphs, R. (2004). The neuroanatomy of emotional memory in humans. In D. Reisberg & P. Hertel (Eds.), *Memory and emotion* (pp. 42–75). New York: Oxford University Press.

Buchanan, T. W., Etzel, J. A., Adolphs, R., & Tranel, D. (2006). The influence of autonomic arousal and semantic relatedness on memory for emotional words. *Int. J. Psychophysiol.*, *61*, 26–33.

Buchanan, T. W., & Lovallo, W. R. (2001). Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology*, *26*, 307–317.

Buchanan, T. W., Tranel, D., & Adolphs, R. (2005). Emotional autobiographical memories in amnesic patients with medial temporal lobe damage. *J. Neurosci.*, *25*, 3151–3160.

Buchanan, T. W., Tranel, D., & Adolphs, R. (2006). Memories for emotional autobiographical events following unilateral damage to medial temporal lobe. *Brain*, *129*, 115–127.

Bush, S. I., & Geer, J. H. (2001). Implicit and explicit memory of neutral, negative emotional, and sexual information. *Arch. Sex. Behav.*, *30*, 615–631.

Cahill, L. (2003). Sex- and hemisphere-related influences on the neurobiology of emotionally influenced memory. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, *27*, 1235–1241.

Cahill, L., Babinsky, R., Markowitsch, H. J., & McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature*, *377*, 295–296.

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., Wu, J., & McGaugh, J. L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proc. Natl. Acad. Sci. USA*, *93*, 8016–8021.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *J. Neurosci.*, *20*, RC99.

Colgrove, F. W. (1899). Individual memories. *Am. J. Psychol.*, *10*, 228–255.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.*, *3*, 201–215.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *J. Exp. Psychol. Gen.*, *13*, 159–180.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *J. Verb. Learn. Verb. Beh.*, *11*, 671–684.

Daselaar, S. M., Rice, H. J., Greenberg, D. L., Cabeza, R., LaBar, K. S., & Rubin, D. C. (2008). The spatiotemporal dynamics of autobiographical memory: Neural correlates of recall, emotional intensity, and reliving. *Cereb. Cortex*, *18*, 217–229.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.*, *16*, 693–700.

Dolan, R. J., Lane, R., Chua, P., & Fletcher, P. (2000). Dissociable temporal lobe activations during emotional episodic memory retrieval. *NeuroImage*, *11*, 203–209.

Dolan, R. J., & Vuilleumier, P. (2003). Amygdala automaticity in emotional processing. *Ann. NY Acad. Sci.*, *985*, 348–355.

Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron*, *42*, 855–863.

Dolcos, F., LaBar, K. S., & Cabeza, R. (2005). Remembering one year later: Role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. *Proc. Natl. Acad. Sci. USA*, *102*, 2626–2631.

Duncan, S., & Barrett, L. F. (2007). The role of the amygdala in visual awareness. *Trends Cogn. Sci.*, *11*, 190–192.

Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with theories of sleep and memory: Sleep, declarative memory, and associative interference. *Curr. Biol.*, *16*, 1290–1294.

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: Passive, permissive, active or none? *Curr. Opin. Neurobiol.*, *16*, 716–722.

Erk, S., Kiefer, M., Grothe, J., Wunderlich, A. P., Spitzer, M., & Walter, H. (2003). Emotional context modulates subsequent memory effect. *NeuroImage*, *18*, 439–447.

Fenker, D. B., Schott, B. H., Richardson-Klavehn, A., Heinze, H. J., & Duzel, E. (2005). Recapitulating emotional context: Activity of amygdala, hippocampus and fusiform cortex during recollection and familiarity. *Eur. J. Neurosci.*, *21*, 1993–1999.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *J. Exp. Psychol. Gen.*, *130*, 681–700.

Frank, M. G., & Benington, J. H. (2006). The role of sleep in memory consolidation and brain plasticity: Dream or reality? *Neuroscientist*, *12*, 477–488.

Frijda, N. H., & Sundararajan, L. (2007). Emotion refinement: A theory inspired by Chinese poetics. *Perspect. Psychol. Sci.*, *2*, 227–241.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learn. Memory*, *13*, 259–262.

Garoff, R. J., Slotnick, S. D., & Schacter, D. L. (2005). The neural origins of specific and general memory: The role of the fusiform cortex. *Neuropsychologia*, *43*, 847–859.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends Cogn. Sci.*, *5*, 394–400.

Hamann, S. (2005). Sex differences in the responses of the human amygdala. *Neuroscientist*, *11*, 288–293.

Hamann, S., & Canli, T. (2004). Individual differences in emotion processing. *Curr. Opin. Neurobiol.*, *14*, 233–238.

Hu, P., Stylos-Allan, M., & Walker, M. P. (2006). Sleep facilitates consolidation of emotionally arousing declarative memory. *Psychol. Sci.*, *10*, 891–898.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspect. Psychol. Sci.*, *2*, 260–280.

Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-neuroanatomic correlates of recollection: Implications for models of recognition memory. *J. Neurosci.*, *24*, 4172–4180.

Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. *Rev. Neurosci.*, *15*, 241–251.

Kensinger, E. A. (2007). How negative emotion affects memory accuracy: Behavioral and neuroimaging evidence. *Curr. Dir. Psychol. Sci.*, *16*, 213–218.

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Mem. Cogn.*, *31*, 1169–1180.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proc. Natl. Acad. Sci. USA*, *101*, 3310–3315.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *J. Mem. Lang.*, *54*, 99–112.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007a). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *J. Mem. Lang.*, *56*, 575–591.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007b). How negative emotion enhances the visual specificity of a memory. *J. Cogn. Neurosci.*, *19*, 1872–1887.

Kensinger, E. A., & Leclerc, C. M. (in press). Age-related changes in the neural mechanisms supporting emotion processing and emotional memory. *Eur. J. Cogn. Psychol.*

Kensinger, E. A., & Schacter, D. L. (2005a). Emotional content and reality-monitoring ability: FMRI evidence for the influence of encoding processes. *Neuropsychologia*, *43*, 1429–1443.

Kensinger, E. A., & Schacter, D. L. (2005b). Retrieving accurate and distorted memories: Neuroimaging evidence for effects of emotion. *NeuroImage*, *27*, 167–177.

Kensinger, E. A., & Schacter, D. L. (2007). Remembering the specific visual details of presented objects: Neuroimaging evidence for effects of emotion. *Neuropsychologia*, *45*, 2951–2962.

Kensinger, E. A., & Schacter, D. L. (2008). Neural processes supporting young and older adults' emotional memories. *J. Cogn. Neurosci.*, *7*, 1–13.

Kern, R. P., Libkuman, T. M., Otoni, H., & Holmes, K. (2005). Emotional stimuli, divided attention, and memory. *Emotion, 5*, 408–417.

Kilpatrick, L., & Cahill, L. (2003). Amygdala modulation of parahippocampal and frontal regions during emotionally influenced memory storage. *NeuroImage*, *20*, 2091–2099.

Kleinsmith, L. J., & Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *J. Exp. Psychol.*, *65*, 190–193.

Koutstaal, W., Wagner, A. D., Rotte, M., Maril, A., Buckner, R. L., & Schacter, D. L. (2001). Perceptual specificity in visual object priming: Functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia*, *39*, 184–199.

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nat. Neurosci. Rev.*, *7*, 54–56.

LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychol. Sci.*, *9*, 490–493.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International Affective Picture System (IAPS): Technical manual and affective ratings*. Gainesville, FL: Center for Research in Psychophysiology.

Leclerc, C. M., & Kensinger, E. A. (2008). Age-related differences in medial prefrontal activation in response to emotional images. *Cogn. Affective Behav. Neurosci., 8*, 153–164.

LeDoux, J. E. (1995). Emotion: Clues from the brain. *Annu. Rev. Psychol.*, *46*, 209–235.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.*, *23*, 155–184.

Maratos, E. J., Allan, K., & Rugg, M. D. (2000). Recognition memory for emotionally negative and neutral words: An ERP study. *Neuropsychologia*, *38*, 1452–1465.

Maratos, E. J., Dolan, R. J., Morris, J. S., Henson, R. N., & Rugg, M. D. (2001). Neural activity associated with episodic memory for emotional context. *Neuropsychologia*, *39*, 910–920.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends Cogn. Sci.*, *11*, 442–450.

Mather, M. (2006). Why memories may become more positive with age. In B. Uttl, N. Ohta, & A. L. Siegenthaler (Eds.), *Memory and emotion: Interdisciplinary perspectives* (pp. 135–159). Malden, MA: Blackwell.

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspect. Psychol. Sci.*, *2*, 33–52.

McDonald, A. J. (2003). Is there an amygdala and how far does it extend? *Ann. NY Acad. Sci.*, *985*, 1–21.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci.*, *27*, 1–28.

Mickley, K. R., & Kensinger, E. A. (2008). Emotional valence influences the neural correlates associated with remembering and knowing. *Cogn. Affective Behav. Neurosci., 8*, 143–152.

Mogg, K., Bradley, B. P., de Bono, J., & Painter, M. (1997). Time course of attentional bias for threat information in nonclinical anxiety. *Behav. Res. Ther.*, *35*, 297–303.

Nader, K. (2003). Memory traces unbound. *Trends Neurosci.*, *26*, 65–72.

Noesselt, T., Driver, J., Heinze, H. J., & Dolan, R. (2005). Asymmetrical activation in the human brain during processing of fearful faces. *Curr. Biol.*, *15*, 424–429.

Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proc. Natl. Acad. Sci. USA*, *97*, 11120–11124.

Ochsner, K. N., & Schacter, D. L. (2003). Remembering emotional events: A social cognitive neuroscience approach. In R. J. Davidson, H. Goldsmith, and K. R. Scherer (Eds.), *Handbook of the affective sciences*. (pp. 643–660). New York: Oxford University Press.

Ohman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *J. Exp. Psychol. Gen.*, *130*, 466–478.

Panksepp, J. (2007). Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspect. Psychol. Sci.*, *2*, 281–296.

Payne, J. D., Ellenbogen, J. M., Walker, M. P., Stickgold, R. (in press). The role of sleep in memory consolidation. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference*. New York: Elsevier.

Payne, J. D., Stickgold, R., Swanberg, K., & Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychol. Sci.*, *19*, 781–788.

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Curr. Opin. Neurobiol.*, *15*, 188–196.

Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Curr. Opin. Neurobiol.*, *14*, 198–202.

Phelps, E. A., LaBar, K. S., Anderson, A. K., O'Connor, K. J., Fulbright, R. K., & Spencer, D. D. (1998). Specifying the contributions of the human amygdala to emotional memory: A case study. *Neurocase*, *4*, 527–540.

Phelps, E. A., LaBar, K. S., & Spencer, D. D. (1997). Memory for emotional words following unilateral temporal lobectomy. *Brain Cogn.*, *35*, 85–109.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*, 175–187.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychol. Sci.*, *17*, 292–299.

Rasch, B., & Born, J. (2008). Maintaining memories by reactivation. *Curr. Opin. Neurobiol.*, *17*, 698–703.

Reisberg, D., & Heuer, F. (2004). Remembering emotional events. In D. Reisberg and P. Hertel (Eds.), *Memory and emotion* (pp. 3–41). New York: Oxford University Press.

Richardson, M. P., Strange, B., & Dolan, R. J. (2004). Encoding of emotional memories depends on the amygdala and hippocampus and their interactions. *Nat. Neurosci.*, *7*, 278–285.

Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Curr. Opin. Neurobiol.*, *6*, 228–236.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends Cogn. Sci.*, *4*, 108–115.

Sakai, K. (2003). Reactivation of memory: Role of medial temporal lobe and prefrontal cortex. *Rev. Neurosci.*, *14*, 241–252.

Scherer, K. R. (2000). Psychological models of emotion. In J. C. Borod (Ed.), *The neuropsychology of emotion* (pp. 137–162). New York: Oxford University Press.

Schultz, W. (2000). Multiple reward signals in the brain. *Nat. Rev. Neurosci.*, *1*, 199–207.

Sergerie, K., Lepage, M., & Armony, J. L. (2006). A process-specific functional dissociation of the amygdala in emotional memory. *J. Cogn. Neurosci.*, *18*, 1359–1367.

Sharot, T., Delgado, M. R., & Phelps, E. A. (2004). How emotion enhances the feeling of remembering. *Nat. Neurosci.*, *7*, 1376–1380.

Sharot, T., Martorella, E. A., Delgado, M. R., & Phelps, E. A. (2007). How personal experience modulates the neural circuitry of memories of September 11. *Proc. Natl. Acad. Sci. USA*, *104*, 389–394.

Sharot, T., Verfaellie, M., & Yonelinas, A. P. (2007). How emotion strengthens the recollective experience: A time-dependent hippocampal process. *PloS One*, *2*, e1068.

Sharot, T., & Yonelinas, A. P. (2008). Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition*, *106*, 538–547.

Smith, A. P., Henson, R. N., Dolan, R. J., & Rugg, M. D. (2004). fMRI correlates of the episodic retrieval of emotional contexts. *NeuroImage*, *22*, 868–878.

Smith, A. P., Henson, R. N., Rugg, M. D., & Dolan, R. J. (2005). Modulation of retrieval processing reflects accuracy of emotional source memory. *Learn. Mem.*, *12*, 472–479.

Smith, A. P., Stephan, K. E., Rugg, M. D., & Dolan, R. J. (2006). Task and content modulate amygdala-hippocampal connectivity in emotional retrieval. *Neuron*, *49*, 631–638.

Smith, C., & Rose, G. M. (1996). Evidence for a paradoxical sleep window for place learning in the Morris water maze. *Physiol. Behav.*, *59*, 93–97.

Somerville, L. H., Wig, G. S., Whalen, P. J., & Kelley, W. M. (2006). Dissociable medial temporal lobe contributions to social memory. *J. Cogn. Neurosci.*, *18*, 1253–1265.

Sterpenich, V., D'Argembeau, A., Desseilles, M., Balteau, E., Albouy, G., Vandewalle, G., Degueldre, C., Luxen, A., Collette, F., & Maquet, P. (2006). The locus ceruleus is involved in the successful retrieval of emotional memories in humans. *J. Neurosci.*, *26*, 7416–7423.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*, 1272–1278.

Tabert, M. H., Borod, J. C., Tang, C. Y., Lange, G., Wei, T. C., Johnson, R., et al. (2001). Differential amygdala activation during emotional decision and recognition memory tasks using unpleasant words: An fMRI study. *Neuropsychologia*, *39*, 556–573.

Talmi, D., Anderson, A. K., Riggs, L., Caplan, J. B., & Moscovitch, M. (2008). Immediate memory consequences of the effect of emotion on attention to pictures. *Learn. Mem.*, *15*, 172–182.

Talmi, D., Luk, B. T. C., McGarry, L. M., & Moscovitch, M. (2007). The contribution of relatedness and distinctiveness to emotionally-enhanced memory. *J. Mem. Lang.*, *56*, 555–574.

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Mem. Cogn.*, *32*, 742–751.

Talmi, D., Schimmack, U., Paterson, T., & Moscovitch, M. (2007). The role of attention and relatedness in emotionally enhanced memory. *Emotion*, *7*, 89–102.

Vaidya, C. J., Zhao, M., Desmond, J. E., & Gabrieli, J. D. (2002). Evidence for cortical encoding specificity in episodic memory: Memory-induced re-activation of picture processing areas. *Neuropsychologia*, *40*, 2136–2143.

Vuilleumier, P., & Driver, J. (2007). Modulation of visual processing by attention and emotion: Windows on causal interactions between human brain regions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *362*, 837–855.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distinct influences of amygdala lesion on visual cortical activation during emotional face processing. *Nat. Neurosci.*, *7*, 1271–1278.

Wagner, U., Gais, S., & Born, J. (2001). Emotional memory formation is enhanced across sleep intervals with high amounts of rapid eye movement sleep. *Learn. Mem.*, *8*, 112–119.

Wagner, U., Hallschmid, M., Rasch, B., & Born, J. (2006). Brief sleep after learning keeps emotional memories alive for years. *Biol. Psychiatry*, *60*, 788–790.

Wagner, U., Kashyap, N., Diekelmann, S., & Born, J. (2007). The impact of post-learning sleep vs. wakefulness on recognition memory for faces with different facial expressions. *Neurobiol. Learn. Mem.*, *97*, 679–697.

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annu. Rev. Psychol.*, *57*, 139–166.

Walker, E. L., & Tarte, R. D. (1963). Memory storage as a function of arousal and time with homogeneous and heterogeneous lists. *J. Verb. Learn. Verb. Behav.*, *2*, 113–119.

Waring, J., Payne, J. D., Schacter, D. L., & Kensinger, E. A. (in press). Impact of individual differences upon emotion-induced memory trade-offs. *Cogn. Emotion*.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proc. Natl. Acad. Sci. USA*, *97*, 11125–11129.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychol. Bull.*, *120*, 3–24.

50 Individual Differences in the Engagement of the Cortex during an Episodic Memory Task

MICHAEL B. MILLER

ABSTRACT There is a wealth of information at the individual level regarding the neural basis of episodic memory that may be lost by relying on group averages. The topography of brain activity underlying an episodic memory task is enormously variable from individual to individual. Despite this variability, individual patterns of brain activity are relatively stable over time. This stability suggests that there are systematic factors, either cognitive or physiological, between individuals that can account for the variability. We have found that individual differences in memory strategy, as well as other factors, can account for a significant portion of the variance between individuals in their patterns of brain activity. These findings demonstrate that, while performance of a typical episodic memory task engages widespread specialized brain regions throughout most of the cortex, different strategies may differentially engage these various brain regions.

It can be just as important to study the things that make us different from each other as it is to study the things that we have in common. This fact has been appreciated since at least 1911, when the eminent learning theorist E. L. Thorndike wrote, "If we could thus adequately describe each of a million human beings, . . . the million men would be found to differ widely. . . . We may study the features of intellect and character which are common to all men; or we may study the differences in intellect and character which distinguish individual men." One of the important things that make us different is the unique ways in which we remember past events. Further, these uniquely individual approaches to memory likely result in extensive variability in the engagement of specialized, universal brain regions. This is particularly evident in the variable pattern of brain activity observed in fMRI studies across individuals performing an episodic memory task. Patterns of individual brain activity are as unique and persistent as fingerprints. Yet, unlike fingerprints, these unique and persistent patterns of brain activity may also be quite informative about individuals and how they go about remembering past events. And, while variability is often treated as a nuisance, controlled by averaging across individuals, relying on a group map can also be a lost opportunity to realize the full extent of the brain's involvement in particular tasks, particularly a task as dynamic, complex, and strategic as episodic memory.

Group maps of whole-brain activity during an episodic retrieval task, no matter how sophisticated and rigorous the statistical analyses, have been shown to be poor representations of the pattern of activations and deactivations that occur at the individual level (Miller et al., 2002; Miller et al., in press). The individual differences in the patterns of activity were observed to be so extensive that one subject, for example, had significant activity in the dorsolateral regions of the prefrontal and parietal cortex while another subject had significant activity in the ventrolateral regions only. Yet, when we brought the subjects back for another session months later, the individual patterns of activity were relatively stable, indicating that a significant portion of the variance between individuals was not due to random fluctuations of noise. We replicated this finding in a recent fMRI study that compared individual patterns of brain activity across repeated sessions (Miller et al., in press). As shown in figure 50.1, the patterns of activations and deactivations were quite unique from individual to individual. Yet the individual patterns of activity persisted over an extended period of time (in this case, between 2 and 4 months). This stability indicated that the variations observed between individuals were not due to random fluctuations, but represented some systematic differences between individuals that were greatly affecting the pattern of brain activity across the whole brain. It has been our observation across several studies of episodic memory that the group maps are not representative of the patterns of activations occurring at the individual level.

The dangers of averaging across subjects

Within the field of psychology, there have been numerous examples over the years of erroneous conclusions based

MICHAEL B. MILLER Department of Psychology, University of California, Santa Barbara, California

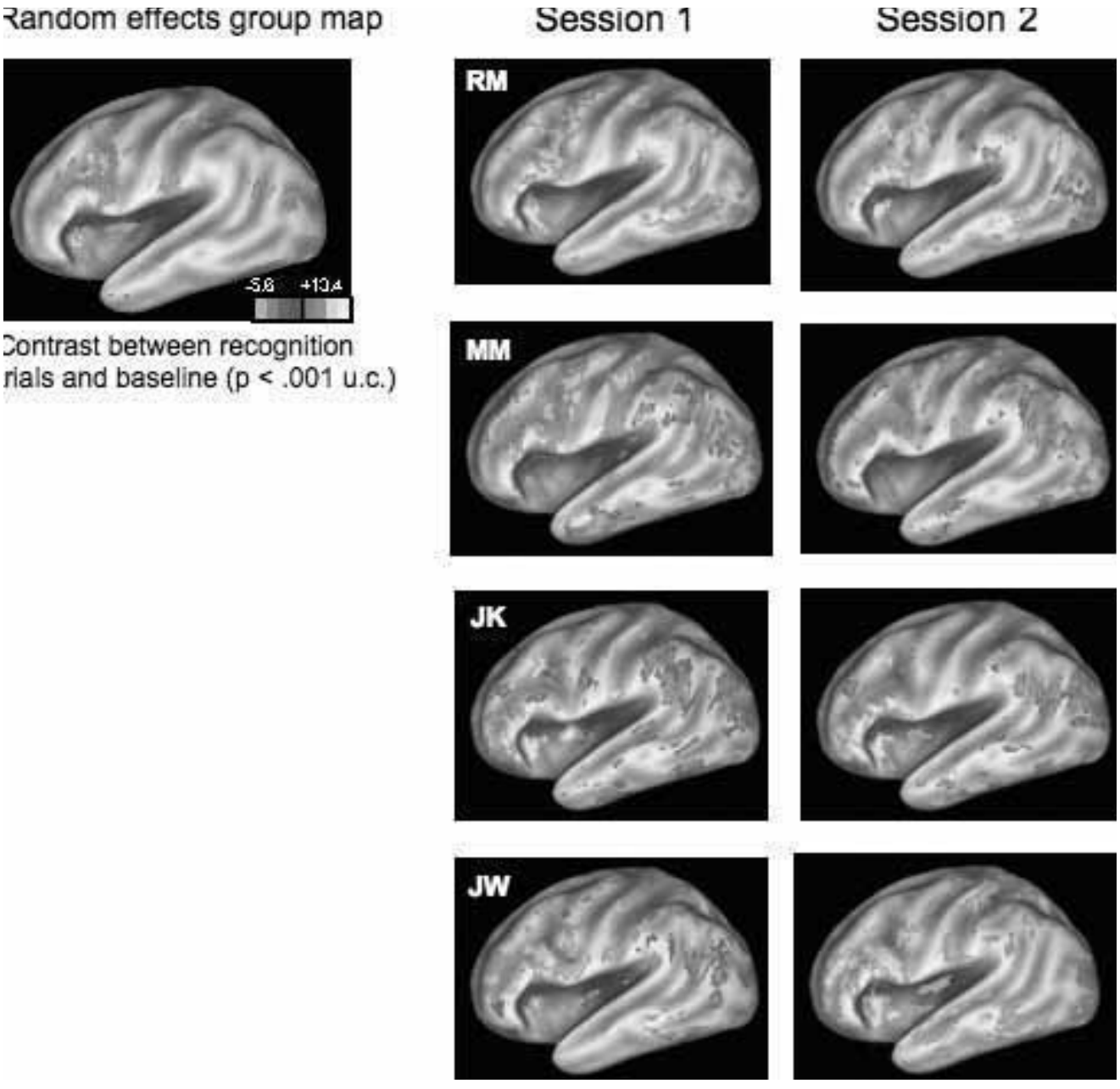


FIGURE 50.1 The left hemisphere patterns of activations and deactivations during an episodic retrieval task for the group and for 4 of the 14 individuals that made up the group. The random-effects group map and the individual *t*-maps are a contrast between the retrieval condition and baseline, statistically thresholded at $p < .001$ uncorrected for multiple comparisons. The thresholding was done for visualization purposes only. Displayed for each individual is the *t*-map for the first and second sessions. The images are representative of the high degree of variability between individuals, yet the relative stability in the individual patterns of activity over time. (From Miller et al., in press.) (See color plate 61.)

on averaged data. In one compelling example, Gallistel, Fairhurst, and Balsam (2004) demonstrated in a basic animal learning paradigm that the negatively accelerated, gradually increasing learning curve (a basic assumption of most learning theorists) is an artifact of averaging across individuals. Using conditioned responses in pigeons (pecking a key), Gallistel and colleagues effectively showed that learning within individual birds was actually abrupt and steplike. An abrupt and steplike learning process is a fundamentally different psychological process from a gradual learning process over time. William Estes recently noted that, whereas a model built on group data can illustrate real trends, it can be a poor fit to individual data, and it can be a major source of distortion. In discussing the efforts of a number of inves-

tigators in the 1950s to raise awareness about the dangers of group data, Estes wrote, "It is not easy, however, to change the habits of people who are comfortable with traditional ways of doing things, and developers of cognitive models have continued to rely for support mainly on the fitting of functions such as curves of learning, retention, and generalization to averaged data" (Estes, 2002, p. 6).

The issue of relying on commonalities across individuals has also been debated repeatedly in the fields of neurology and neuropsychology for over a century (Caramazza, 1986; Sokol, McCloskey, Cohen, & Aliminosa, 1991; Robertson, Knight, Rafal, & Shimamura, 1993). In the mid-1800s, Paul Broca argued that speech production could be localized to the third convolution of the left inferior frontal gyrus by examining the common area of damage across a group of patients exhibiting similar speech production deficits. At the same time, however, another neurologist, John Hughlings Jackson, argued against a centralized region for speech based on his observations of wide variations in the extent and location of damage in patients exhibiting similar aphasic symptoms and wide variations in symptoms in patients with similar damage. Contrary to Broca, Jackson proposed that speech was a widely distributed function within the brain (Critchley & Critchley, 1998). Of course, Broca's view held sway for the next century and eventually led to models of language, such as the Wernicke-Geschwind model, that relied on these distinct and localized modules of language function. However, more recent studies have suggested that, although this classic brain-language model is useful as a heuristic, it is empirically wrong because it cannot account for the range of aphasic symptoms, it is underspecified linguistically and anatomically, and it does not take into account the extensive individual variability in symptoms of brain-damaged patients and in the location and extent of their damage (Benson, 1985; Poeppel & Hickok, 2004). For example, Nina Dronkers (1996) reported on 22 patients with lesions in Broca's area with only 10 having Broca's aphasia. Ojemann, Ojemann, Lettich, and Berger (2008) reported on cortical stimulation during neurosurgery of 117 patients and found that language disruption occurred in individualized mosaics of cortex with substantial variability in the location of these mosaics, some of which correlated with sex and intelligence. It is clear that, on one hand, localizing brain functions by isolating common areas of brain damage across patients with similar deficits has been useful and, to a certain extent, necessary given that brain damage is rarely confined to one specific functional region, and that reliance on case studies can "sacrifice generalizability, predictability, and the possibility of refutation" due to subject variability (Robertson et al., 1993, p. 716). On the other hand, case-study-by-case-study approaches can accomplish a similar winnowing and modeling of function/brain relationships while preserving much of the information that is at the individual level but lost in a group average (Caramazza, 1986; Sokol et al., 1991).

Individual differences and neuroimaging

Neuroimaging faced a similar issue in its early years. Marc Raichle has commented that many of the early researchers worried that the creation of group maps by averaging neuroimaging data across subjects would greatly diminish the signal due to inherently high individual variability (Raichle, 1997). Yet those early studies reliably demonstrated retinotopic mapping of the primary visual cortex (Fox et al., 1986), as well as mapping of higher-order association areas (Petersen, Fox, Posner, Mintun, & Raichle, 1988), using group maps. Since that time, group maps have become much more sophisticated and population-based, and more emphasis has been placed on them given the struggle to overcome the inherently low overall signal-to-noise ratio of neuroimaging data. It is interesting to note, however, that many vision researchers have reverted back to relying on individual data by retinotopically mapping individuals and testing hypotheses on an individual basis with many trials (Warnking et al., 2002). In addition, many researchers are relying on functional localizers in several tasks because of individual differences in the specific location of specialized regions within the brain (Saxe, Brett, & Kanwisher, 2006). We argue that a general reliance on group maps to represent the pattern of activity across the whole brain for a particular task needs to be reevaluated.

Few studies have attempted to systematically examine the individual variability of brain activity associated with a cognitive task. In general, most neuroimaging studies involving individual differences can be divided into four categories: (1) studies that correlate a particular behavioral performance with modulated activity in a specific brain region; (2) studies that divide subjects into smaller groups based on a behavioral measure and then look for differences in activations between the groups; (3) studies that look at the overlap of individual brain activations and variations of activity around a circumscribed region; and (4) studies that look at the degree to which group activations are reproducible. Each of these techniques has been reviewed previously (Miller & Van Horn, 2007), and each can be a useful analytical tool. For example, a convincing way to demonstrate the function of a given brain region is to show that the activity in that region is modulated by individual differences in behavior, as has been demonstrated in numerous neuroimaging studies from correlations of individual differences in procedural learning and the modulation of activity in the motor cortex (Grafton, Woods, & Tyszka, 1994) to individual differences in memory performance and modulation of the medial temporal lobe (Nyberg, McIntosh, Houle, Nilsson, & Tulving, 1996; Tulving, Habib, Nyberg, Lepage, & McIntosh, 1999). Some

of these studies have even considered the correlation between individual differences in memory-encoding strategies and particular brain regions or differences in the patterns of activity by grouping individuals by particular encoding strategies (Savage et al., 2001; Casasanto et al., 2002; Kirchhoff & Buckner, 2006). Yet most of these studies still rely on a common area of activation across a group of subjects, and only a few studies to our knowledge consider the individual variability and reliability of activity across the whole brain volume (Heun et al., 2000; Machielsen, Rombouts, Barkhof, Scheltens, & Witter, 2000; Miller et al., 2002).

One notable exception was a study conducted by McGonigle and colleagues (2000). Although the study consisted of only a single subject, that subject participated in 33 scanning sessions and performed a simple motor, visual, and cognitive task in each session. The authors found that many voxels displaying significant session-by-condition interactions were not activated on average across all the sessions, and many of those that were activated using a fixed-effects analysis did not survive a random-effects analysis across sessions. The authors cautioned that single-session data from individuals may lead to erroneous conclusions. But a subsequent follow-up study by the same authors revealed that the intersession variability was no greater than the intrasession variability (Smith et al., 2005).

The effectiveness or ineffectiveness of group maps may depend on the intended use. For instance, group maps can be very effective when used as a tool to increase the signal-to-noise ratio in a study that is systematically examining some a priori brain region. Wagner and colleagues (1998), for example, utilized a group map to examine the BOLD response in the medial temporal lobe. They demonstrated that activity in the medial temporal lobe during an encoding task varied according to whether an item was subsequently recognized. This was a critically important study given that activation in the medial temporal lobe was conspicuously absent in most previous neuroimaging studies of episodic memory. Group maps used in this way can only aid our understanding of mind/brain relationships.

However, group maps can be a problem if they are meant to characterize or profile the pattern of activity across the whole brain for a given task. For example, we conducted a study with the purpose of characterizing the pattern of activity associated with shifts of criterion on a recognition test (Miller, Handy, Cutler, Inati, & Wolford, 2001). We published several group maps, including one meant to characterize activity during blocks of recognition trials compared to blocks of fixation trials. The results of a group analysis revealed activations associated with recognition in discreet regions of the dorsolateral prefrontal and parietal cortex. However, a subsequent analysis revealed that the group map was not representative of a pattern of activations in many of the individual subjects (Miller et al., 2002). In the end, using group maps to profile patterns of activity may be more appropriate for some cognitive tasks and not others. For example, some cognitive tasks may engage a more restricted set of brain regions that vary little from subject to subject than does an episodic retrieval task. In general, though, we think caution should be used when employing a group map to characterize patterns of activity across the whole brain for a given task. Does a group map represent the patterns of activations for the individuals that make up the group? For episodic retrieval tasks, the answer is no.

Quantifying the degree of similarity between two patterns of brain activity

It is clear from viewing the thresholded activation maps shown in figure 50.1 that there is enormous variability from individual to individual, but we needed a method to quantify those differences. Furthermore, we needed a method that did not rely on the arbitrary setting of a statistical threshold. So we devised a method to simply cross-correlate the unthresholded image volumes across subjects and sessions (Miller et al., 2002). If one takes a three-dimensional volume of continuous values in each voxel and correlates that volume with another three-dimensional volume of continuous values in the same voxel matrix and atlas space, then the result will be a single correlation value that represents how similar the two volumes are. A similar correlational approach has been used with smaller patches of cortex in pattern classification studies of object recognition (Haxby, 2004; Norman, Polyn, Detre, & Haxby, 2006). In our original study (Miller et al., 2002) we correlated volumes of raw signal intensity values, but those correlations may have been sensitive to individual differences in basic physiology, such as individual differences in vasculature and individual differences in the timing of the BOLD response. In order to attenuate those particular differences in physiology in recent studies, we correlated volumes containing unthresholded *t*-values (*t*-maps) derived from the contrast between task trials and baseline within each individual (Miller et al., in press) and masked (with masks derived from the group analysis) to exclude extra-brain voxels. This method provides a convenient measure of the degree of similarity between any two volumes of brain activity.

Using this method, we are able to quantify the observation that individual brain activity during the episodic retrieval task is quite variable from individual to individual, yet it is also relatively stable over time. In one of our recent studies, we found that the average correlation between two volumes from the same subject performing the same episodic retrieval task but in different sessions separated by 2 to 4 months was $r = .435$. This was significantly higher than the average correlation between two volumes from different subjects performing the same task, which was $r = .218$ (Miller et al.,

in press). The relative stability of the individual patterns of activity over long periods of time suggested that unique individual activations were not necessarily noise but instead were likely to reflect cognitive processing that was unique to the individual and was related to how the individual performed the task and/or other unique physiological properties related to that individual. The persistence of these uniquely individual patterns of activity should be viewed as an opportunity and not as a nuisance. The opportunity it affords us is the ability to explore the fundamentally different ways we remember past events and the unique brain regions we recruit to accomplish that task.

The inherently variable nature of episodic memory and the brain regions underlying it

Episodic memory "stores and makes possible subsequent recovery of information about personal experiences from the past. It enables people to travel back in time, as it were, into their personal past, and to become consciously aware of having witnessed or participated in events and happenings at earlier times." (Tulving, 1989, p. 362). However, the methods used to probe episodic memory experimentally, such as a standard recognition test, utilize not only bits of information from episodic memory but bits of information from other systems as well, such as semantic memory. As Endel Tulving once wrote, "It is probably as difficult to find 'pure' episodic-memory tasks and 'pure' semantic tasks as it is to find sodium and chlorine as free elements in nature, although their compound, NaCl, is found in abundance" (Tulving, 1983, p. 55). A "remembered" response on a recognition test can be influenced by a variety of nonepisodic processes like semantic associations (Underwood, 1965), schematic reconstructions (Brewer & Treyens, 1981; Miller & Gazzaniga, 1998), perceptual fluency (Jacoby & Dallas, 1981), shifting criterion (Miller & Wolford, 1999), and so on (for reviews see Roediger, 1996; Schacter, 1999).

Further, episodic memory is widely distributed throughout the cortex, with different regions storing different aspects of the complete memory trace (Squire, 1987). It relies on an extensive hippocampal-cortical network for the consolidation, storage, and utilization of information, and the hippocampus is not involved in the permanent storage of information per se, but rather serves to facilitate consolidation of a distributed cortical memory trace (Squire et al., 1992; Wittenberg & Tsien, 2002; but see also Nadel & Moscovitch, 1997). A principal characteristic of this distributed network is that it affords the rapid and flexible formation of multimodal memories. In addition to widely distributed information, there is also a broad network of specialized brain regions that are influential in, but not necessary for, the completion of the task. After all, only damage to the medial temporal lobe causes severe amnesia (Squire, Stark, & Clark, 2004; Eichenbaum, Yonelinas, & Ranganath, 2007). Many memory researchers have suggested that prefrontal and parietal areas support episodic memory with cognitive processes peripheral to the actual retrieval process, with evidence derived from brain-damaged patient studies (Incisa della Rocchetta & Milner, 1993; Janowsky, Shimamura, Kritchevsky, & Squire, 1989; Petrides, 1996; Ranganath, Johnson, & D'Esposito, 2003; Knight, 1991) and neuroimaging studies (Nyberg et al., 1995; Buckner, Koustaal, Schacter, Wagner, & Rosen, 1998; Rugg et al., 1998; Fletcher, Shallice, Frith, Frackowiak, & Dolan, 1998; Cabeza et al., 2003; Nolde, Johnson, & D'Esposito, 1998; Henson, Shallice, & Dolan, 1999; Dobbins, Rice, Wagner, & Schacter, 2003). One potential implication of this architecture is that one and the same behavioral outcome—such as an "old" response on a recognition test—could be based on a distinct set of information and a distinct combination of neural circuits in two different individuals.

Therefore the emerging picture of the neural basis of episodic retrieval is that it comprises several distinct brain regions and that these distinct brain regions may be engaged differentially depending on unique individual strategies and demands. There is substantial evidence that people will employ a multitude of strategies during the encoding and retrieval phases of a standard memory task (Stoff & Eagle, 1971; Battig, 1975; Weinstein, Underwood, Wicker, & Cubberly, 1979; Paivio, 1983; Reder, 1987; Graf & Birt, 1996). There is also substantial evidence from neuroimaging studies that individual differences in memory strategy can alter which brain regions become activated (Savage et al., 2001; Casasanto et al., 2002; Speer, Jacoby, & Braver, 2003; Kondo et al., 2005; Tsukiura, Mochizuki-Kawai, & Fujii, 2005). One notable study (Kirchoff & Buckner, 2006) identified the various strategies people adopt during an unconstrained encoding of unrelated pairs of pictures. They found that two strategies in particular, verbal elaboration and visual inspection, correlated with memory performance and with brain activity in distinct regions: verbal elaboration correlated with activity in prefrontal regions associated with controlled verbal processing, whereas visual inspection correlated with activity in the extrastriate cortex.

The variable, unconstrained, and widely distributed nature of brain activity during an episodic retrieval task is particularly evident in the reported sites of activations across studies when compared to the reported sites of activations from other tasks, such as semantic retrieval. Cabeza and Nyberg (2000) categorized hundreds of neuroimaging studies by cognitive domain and then plotted the reported sites of activations for each study within each cognitive domain as a point on a glass brain. A cursory review of their findings reveals a general consistency in the localization of activity across studies for most of the cognitive domains, but not for episodic retrieval. Even in more constrained versions of

the task, the reported sites of activation were widely distributed throughout the cortex. However, the reported sites of activations for other tasks appear to be more concentrated in discrete regions of the cortex. This pattern may result from cognitive processing that is more constrained in general during a controlled semantic retrieval task, for example, than during an episodic retrieval task. Controlled semantic retrieval is known to unambiguously engage the left ventrolateral prefrontal cortex (Wagner, Pare-Blagoev, Clark, & Poldrack, 2001). For this task, the divergence of brain regions that any particular individual is likely to engage in order to accomplish the task may be relatively small.

We recently found that the topographical pattern of brain activity during an episodic retrieval task (subjects responded "old" or "new" to previously studied words) was significantly more variable between individuals than a semantic retrieval task (subjects responded "abstract" or "concrete" to words similar to the episodic task) or a working memory task (subjects responded "match to three trials back" or "no match" to a sequence of letters) (Miller et al., 2002). Further, by examining the variance across individuals on a voxel-by-voxel basis, we found that the variance in activity during an episodic memory task occurred throughout the whole cortex, whereas the variance in activity during a semantic memory task and a working memory task occurred in much more discrete regions of the cortex (see figure 50.2). However, some of the differences between tasks could be attributed to different demand characteristics. Also, there is no reason to assume that a task like working memory could not have some of the same properties (the variable and strategic engagement of widely distributed brain regions) as an episodic memory task. Future research will need to determine the extent to which brain activity during an episodic memory task may be more variable than other cognitive tasks.

What makes us unique?

The opportunity afforded by an inherently strategic task like episodic retrieval and the brain activity underlying it is that it provides a lot of useful information at the individual level that may enhance our understanding of episodic memory in general. Our previous results indicate that a large portion of the variance in brain activity from individual to individual is not necessarily random and that some factor or factors must be accounting for the variance systematically (Miller et al., 2002, in press). In order to understand the sources of this variability, we have recently focused on three general categories of factors: (1) situational differences; (2) anatomical/physiological differences, and (3) cognitive/psychological differences.

In a recent study (Miller et al., in press), we investigated some of these possible sources of individual variability during an episodic retrieval task. During the episodic retrieval task, subjects simply made an "old/new" recognition judgment to words, some of which were previously studied. Fourteen subjects were scanned in two sessions separated by 2 to 4 months. After cross-correlating the *t*-maps from a recognition versus baseline contrast across subjects and sessions, the correlation values were then submitted to a hierarchical regression analysis in order to determine whether certain factors could account for the degree of similarity between subjects. In this study, we had a limited number of factors, but they included anatomical similarity, memory performance (d'), reaction time, and retrieval strategy (as measured by a decision criterion). The placement of a decision criterion during a recognition test is often the result of a general strategy or bias (Murdock, 1974; Ratcliff, Sheu, & Gronlund, 1992; Miller & Wolford, 1999). For example, some subjects may have responded "old" to a test item only if they were absolutely certain they encountered the item during the study session (maybe based on some clear visual recollection). The criterion measures from those subjects would have tended to be conservative. For example, in a debriefing after the scanning session, one subject stated, "I only said 'recognize' when I was pretty certain," and her criterion measure reflected a conservative strategy ($C = +.46$). Other subjects may have simply responded "old" to any item that seemed familiar to them regardless of whether or not they had a clear recollection of the item in the study session. For example, another subject stated, "The recognition test was kind of hard, but I would just press 'yes' if the words seemed familiar." Her criterion measure ($C = -.23$) reflected a much more liberal strategy, one based more on familiarity. One of the key findings of this study was that only the difference in criterion values between two subjects was predictive of the similarity in their patterns of brain activity. No other factor was predictive. Therefore, the more similar two individuals' retrieval strategy was, the more similar was their pattern of brain activity during an episodic memory task.

A number of encoding and retrieval strategies have been identified, decision criterion placement being just one example. The most extensively studied strategy, however, has been elaboration, including both imaginal elaboration (mental imagery) and verbal elaboration (Paivio, 1971). For example, some people (who could be called verbalizers) are better at processing words and may rely on semantic associations and verbal content when remembering a past event, while other people (who could be called visualizers) are better at processing pictures and may rely on visual imagery and visual recollections when remembering a past event. To test whether the tendency or preference of a person to think visually or verbally is related to the observed variability in an episodic memory task, we conducted a study in which 20 subjects studied lists of highly imageable words, which allow for both verbal and visual elaboration (Donovan & Miller, 2008). Participants were simply

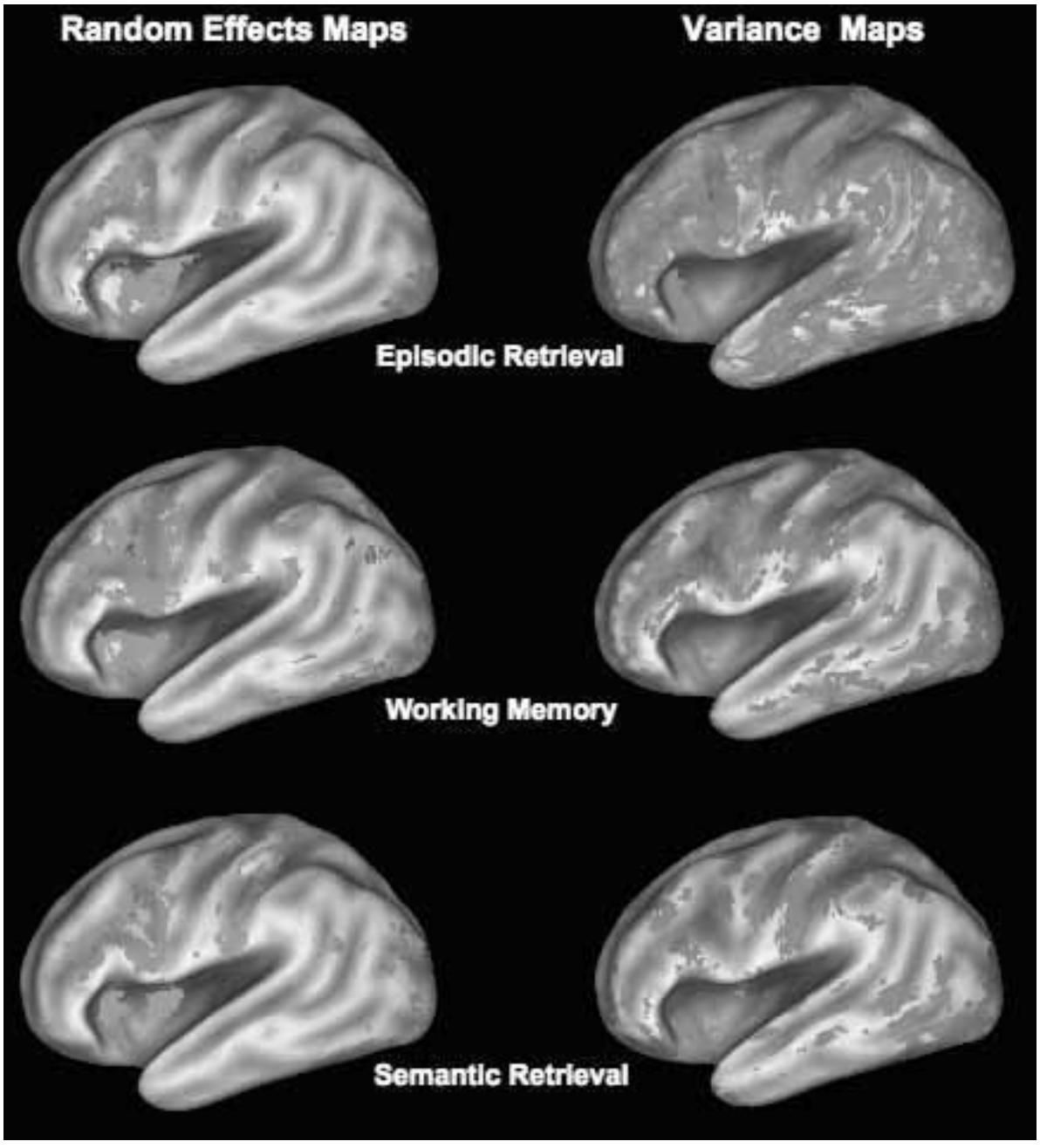


FIGURE 50.2 A comparison of random-effects group maps and variance maps across three memory tasks. The random-effects maps are a statistically thresholded ($p < .001$ uncorrected for multiple comparisons) representation of the common areas of brain activity across 14 individuals. The variance maps to the right display the standard deviations across the 14 individuals at each voxel above a threshold of 2 standard deviations. As the variance maps indicate, individuals variably engaged much wider regions of the cortex during episodic retrieval than during semantic retrieval or working memory. (See color plate 62.)

instructed to learn the words for a later recognition memory test and hence were free to choose whatever strategy came most naturally. During the episodic retrieval task, subjects simply made an "old/new" recognition judgment of the words, half of which were previously studied. In a hierarchical regression analysis, we included several factors to assess their relative contribution to the observed variability: anatomical similarity, connectivity similarity (measured by computing fractional anisotropy maps from DTI images), default mode network similarity, encoding strategy, visualizer/verbalizer trait factor scores, and performance measures. As predicted, we found that the more similar two individuals' tendency to visualize, the more similar their patterns of brain activity.

In this same study, we also assessed whether differences in the default mode network between individuals might also be related to the similarity of their activity patterns during an episodic memory task. We computed functional connectivity for three seed regions from functional data collected while the subjects were at rest. Coherence maps were constructed by computing the coherence of the time series of each voxel with that of seed regions in the posterior cingulate/precuneus (PCC), the ventral anterior cingulate cortex (vACC) (Greicius, Krasnow, Reiss, & Menon, 2003), and the hippocampal formation (Vincent et al., 2006). Interestingly, we found that similarity in coherence maps predicted similarity in functional activity during retrieval but not during encoding. Schacter, Addis, and Buckner (2008) have made the suggestion that the default mode network may serve as a simulator of past and future events, which would explain why the relationship between coherence maps and functional activity occurs during the retrieval task but not during the encoding task.

Table 50.1 lists some of the factors that have been found across several studies to be related to the degree of similarity between any two individuals' brain activity patterns. In terms of situational factors, circumstances of the experimental setup can play a major role in the degree of similarity in brain activity between individuals. Two of the strongest factors are the experimental design and the stimulus type. We have found that the difference between a blocked design and event-related design can account for 40% of the variance in correlation values, with brain activity from a blocked design being much less variable between individuals than an event-related design. We have also found that stimulus type can account for 28% of the variance, with brain activity associated with remembering faces being much less variable between individuals than brain activity associated with remembering words. These are remarkably strong factors given that the difference between memory tasks discussed earlier accounted for only 8% of the variance. Another key situational factor is constraining the task. We recently found that making the recollection of a studied item easy to recall resulted in much less variability between individuals than when the recollection of a studied item is more difficult. All of these situational factors must be carefully controlled or accounted for when considering the influence of individual difference factors.

TABLE 50.1

Factors that are related to the degree of similarity between any two brain volumes of activity during an episodic memory task

Factors Related to Variability in Brain Activity	ΔR^2
Situational Factors	
Experimental design: blocked or event-related	40%
Stimulus type (faces or words)	28%
Different sessions	n.s.
Different tasks	8%
Task difficulty	5%
Individual Differences in Physiology and Anatomy	
Structural anatomy	n.s.
Default mode network (coherence maps)	4%
White matter connectivity (fractional anisotropy)	7%
Individual Differences in Cognition and Information Processing	
Retrieval strategy (criteria)	8%
Memory performance (d prime)	n.s.
Reaction time	n.s.
Tendency to visualize	5%
Tendency to verbalize	n.s.
Individual Deviations Unaccounted For	**16–44%**

Data from Miller et al., in press; Donovan & Miller, 2008; Guerin & Miller, 2009. ΔR^2 values are from hierarchical regression analyses conducted in each study with the variables entered in the order noted on the table. Not all variables were represented in each study. The values varied considerably from study to study, with representative values listed here. Factors with "n.s." were not significant in any of the studies.

In terms of individual differences in physiology and anatomy, we have found that similarity in white matter connectivity and in the default mode network are both factors, but not individual differences in structural anatomy. Although all brain volumes are spatially normalized before being analyzed, there still exists a considerable difference between individuals in the orientation and precise location of cortical landmarks. Yet those anatomical differences are not predictive of functional activation differences. As for individual differences in cognition and information processing, it is becoming clear in our studies using episodic memory tasks that individual differences in strategy have a significant effect, but not individual differences in memory performance and accuracy. Much work still needs to be done to determine the full range of factors that contribute to the individual variability. This need is evident in the last factor that is listed in table 50.1. In all our hierarchical regression analyses we include dummy variables for each individual. These dummy variables are always entered last after accounting for all the other individual difference factors. Yet these individual variables still account for around 40% of the variance, suggesting that some individuals are more deviant from the group than others and that we have yet to find the factors that account for that fact.

It should also be noted that many of the factors that distinguish individuals on an episodic memory task may also distinguish those individuals on other tasks as well. We found in the study comparing activity across three different memory tasks (an episodic retrieval task, a working memory task, and a semantic retrieval task) that the brain activity of an individual performing an episodic retrieval task is more similar on average to that same individual performing an entirely different task than it is to a different individual performing the same episodic retrieval task (Miller et al., in press).

Conclusion

Individuals vary enormously in their patterns of brain activity. This variability is particularly widespread during an episodic memory task, and it extends across most of the cortex. These fluctuations in activity between individuals are not random because they are relatively stable over time. Furthermore, we have been able to account for significant portions of this variability, including differences in white matter connectivity and differences in strategy. A clear and full understanding of all the sources of variability between individuals will be necessary in order for us to determine whether a pattern of brain activity observed during a memory task is truly reflective of that individual's thoughts and traits. It will be critical for future neuroimaging studies of episodic memory to explore what makes us unique as well as to explore what we have in common.

ACKNOWLEDGMENTS The author would like to acknowledge that the research discussed in this chapter was supportorted by the Institute for Collaborative Biotechnologies through contract no. W911NF-07-1-0072 from the U.S. Army Research Office.

REFERENCES

Battig, W. (1975). Within-individual differences in "cognitive" processes. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 195–228). Hillsdale, NJ: Erlbaum.

Benson, D. F. (1985). Aphasia and related disorders: A clinical approach. In M.-M. Mesulam (Ed.), *Principles of behavioral neurology* (pp. 193–238). Philadelphia: Davis.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cogn. Psych.*, *13*, 207–230.

Buckner, R. L., Koustaal, W., Schacter, D. L., Wagner, A. D., & Rosen, B. R. (1998). Functional–anatomic study of episodic retrieval using fMRI. I. Retrieval effort vs. retrieval success. *NeuroImage*, *7*, 151–162.

Cabeza, R., Dolcos, F., Prince, S. E., Rice, H. J., Weissman, D. H., & Nyberg, L. (2003). Attention-related activity during episodic memory retrieval: A crossfunction fMRI study. *Neuropsychologia*, *41*, 390–399.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition. II. An empirical review of 275 PET and fMRI studies. *J. Cogn. Neurosci.*, *12*, 1–47.

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain Cogn.*, *5*, 41–66.

Casasanto, D. J., Killgore, W. D. S., Maldjian, J. A., Glosser, G., Alsop, D. C., Cooke, A. M., Grossman, M., & Detre, J. A. (2002). Neural correlates of successful and unsuccessful verbal memory encoding. *Brain Lang.*, *80*, 287–295.

Critchley, M., & Critchley, E. A. (1998). *John Hughlings Jackson: Father of English neurology*. New York: Oxford University Press.

Dobbins, I. G., Rice, H. J., Wagner, A. D., & Schacter, D. L. (2003). Memory orientation and success: Separable neurocognitive components underlying episodic recognition. *Neuropsychologia*, *41*, 318–333.

Donovan, C. L., & Miller, M. B. (2008). Individual variability in brain activity during episodic encoding and retrieval: How it relates to anatomy, strategy, visual/verbal traits and personality. *Soc. Neurosci. Abstracts*, Washington, DC.

Dronkers, N. F. (1996). A new brain region for speech: The insula and articulatory planning. *Nature*, *384*, 159–161.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.*, *30*, 123–152.

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychon. Bull. & Rev.*, *9*(1), 3–25.

Fletcher, P. C., Shallice, T., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1998). The functional roles of prefrontal cortex in episodic memory. II. Retrieval. *Brain*, *121*, 1249–1256.

Fox, P. T., Mintun, M. A., Raichle, M. E., Miezin, F. M., Allman, J. M., & Van Essen, D. C. (1986). Mapping human visual cortex with positron emission tomography. *Nature*, *323*, 806–809.

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proc. Natl. Acad. Sci. USA*, *101*(36), 13124–13131.

Graf, P., & Birt, A. R. (1996). Explicit and implicit memory retrieval: Intentions and strategies. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 25–44). Mahwah, NJ: Erlbaum.

Grafton, S. T., Woods, R. P., & Tyszka, J. M. (1994). Functional imaging of procedural motor learning: Relating cerebral blood flow with individual subject performance. *Hum. Brain Mapping*, *1*, 221–234.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. USA*, *100*, 253–258.

Guerin, S. A., & Miller, M. B. (2009). Lateralization of the parietal old/new effect: An event-related fMRI study comparing recognition memory for words and faces. *NeuroImage*, *44*(1), 232–242.

Haxby, J. V. (2004). Analysis of topographically organized patterns of response in fMRI data: Distributed representations of objects in ventral temporal cortex. In N. Kanwisher & J. Duncan (Eds.), *Attention and performance* (pp. 83–98). New York: Oxford University Press.

Henson, R. N., Shallice, T., & Dolan, R. J. (1999). Right prefrontal cortex and episodic memory retrieval: A functional MRI test of the monitoring hypothesis. *Brain*, *122*, 1367–1381.

Heun, R., Jessen, F., Klose, U., Erb, M., Granath, D. O., Freymann, N., & Grodd, W. (2000). Interindividual variation of cerebral activation during encoding and retrieval of words. *Eur. Psychiatry*, *15*, 470–479.

Incisa della Rocchetta, A., & Milner, B., (1993). Strategic search and retrieval inhibition: The role of the frontal lobes. *Neuropsychologia*, *31*, 503–524.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *J. Exp. Psychol. Gen., 3*, 306–340.

Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behav. Neurosci.*, *103*(3), 548–560.

Kirchhoff, B. A., & Buckner, R. L. (2006). Functional-anatomic correlates of individual differences in memory. *Neuron*, *51*, 263–274.

Knight, R. T. (1991). Evoked potential studies of attention capacity in human frontal lobe lesions. In H. S. Levin, H. M. Eisenberg, et al. (Eds.), *Frontal lobe function and dysfunction* (pp. 139–153). New York: Oxford University Press.

Kondo, Y., Suzuki, M., Mugikura, S., Abe, N., Takahashi, S., Iijima, T., & Fujii, T. (2005). Changes in brain activation associated with use of a memory strategy: A functional MRI study. *NeuroImage*, *24*, 1154–1163.

Machielsen, W. C. M., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., & Witter, M. P. (2000). fMRI of visual encoding: Reproducibility of activation. *Hum. Brain Mapping*, *9*,156–164.

McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R. S. J., & Holmes, A. P. (2000). Variability in fMRI: An examination of intersession differences. *NeuroImage*, *11*, 708–734.

Miller, M. B., Donovan, C. L., Van Horn, J. D., German, E., Sokol-Hessner, P., & Wolford, G. L. (in press). Unique and persistent individual patterns of brain activity across different memory retrieval tasks. *NeuroImage*.

Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, *36*(6), 513–520.

Miller, M. B., Handy, T. C., Cutler, J., Inati, S., & Wolford, G. L. (2001). Brain activations associated with shifts in response criterion on a recognition test. *Can. J. Exp. Psychol.*: Special Issue: *Cognitive Neuroscience*, *55*, 164–175.

Miller, M. B., & Van Horn, J. D. (2007). Individual variability in brain activations associated with episodic retrieval: A role for large-scale databases. *Int. J. Psychophysiol.*, *63*, 205–213.

Miller, M. B., Van Horn, J., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., Grafton, S., & Gazzaniga, M. S. (2002). Extensive individual differences in brain activations during episodic retrieval are reliable over time. *J. Cogn. Neurosci.*, *14*, 1200–1214.

Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychol. Rev.*, *106*(2), 398–405.

Murdock, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr. Opin. Neurobiol.*, *7*, 217–227.

Nolde, S. F., Johnson, M. K., & D'Esposito, M. (1998). Left prefrontal activation during episodic remembering: An event-related fMRI study. *NeuroReport*, *9*, 3509–3514.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.*, *10*(9), 424–430.

Nyberg, L., McIntosh, A. R., Houle, S., Nilsson, L. G., & Tulving, E. (1996). Activation of medial temporal structures during episodic memory retrieval. *Nature*, *380*, 715–717.

Nyberg, L., Tulving, E., Habib, R., Nilsson, L. G., Kapur, S., Houle, S., Cabeza, R., & McIntosh, A. R. (1995). Functional brain maps of retrieval mode and recovery of episodic information. *NeuroReport*, *7*, 249–252.

Ojemann, G., Ojemann, J., Lettich, E., & Berger, M. (2008). Cortical language localization in left, dominant hemisphere: An electrical stimulation mapping investigation in 117 patients. *J. Neurosurg.*, *108*(2), 411–421.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.

Paivio, A. (1983). The empirical case for dual coding. In J. C. Yuille (Ed.), *Imagery, memory, and cognition: Essays in honor of Allan Paivio* (pp. 39–63). Hillsdale, NJ: Erlbaum.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*, 585–589.

Petrides, M. (1996). Lateral frontal cortical contribution to memory. *Sem. Neurosci.*, *8*(1), 57–63.

Poeppel, D., & Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition*, *92*, 1–12.

Raichle, M. E. (1997). Brain imaging. In M. S. Gazzaniga (Ed.), *Conversations in the cognitive neurosciences* (pp. 15–33). Cambridge, MA: MIT Press.

Ranganath, C., Johnson, M. K., & D'Esposito, M. D. (2003). Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia*, *41*, 378–389.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global models of memory using ROC curves. *Psychol. Rev.*, *99*, 518–535.

Reder, L. M. (1987). Beyond associations: Strategic components in memory retrieval. In D. S. Gorfein, & R. R. Hoffman (Eds.),

Memory and learning: The Ebbinghaus Centennial Conference (pp. 203–220). Hillsdale, NJ: Lawrence Erlbaum.

Robertson, L. C., Knight, R. T., Rafal, R., & Shimamura, A. P. (1993). Cognitive neuropsychology is more than single-case studies. *J. Exp. Psychol. Learn. Mem. Cogn.*, *19*, 710–717.

Roediger, H. L., III (1996). Memory illusions. *J. Mem. Lang.*, *35*, 76–100.

Rugg, M. D., Fletcher, P. C., Allan, K., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1998). Neural correlates of memory retrieval during recognition memory and cued recall. *NeuroReport*, *8*, 262–273.

Savage, C. R., Deckersbach, T., Heckers, S., Wagner, A. D., Schacter, D. L., Alpert, N. M., Fischman, A. J., & Rauch, S. L. (2001). Prefrontal regions supporting spontaneous and directed application of verbal learning strategies: Evidence from PET. *Brain*, *124*, 219–231.

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30*(4), 1088–1096.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *Am. Psychol.*, *54*, 182–203.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2008). Episodic simulation of future events: Concepts, data, and applications. *The Year in Cognitive Neuroscience 2008. Ann. NY Acad. Sci.*, *1124*, 39–60.

Smith, S. M., Beckmann, C. F., Ramnani, N., Woolrich, M. W., Bannister, P. R., Jenkinson, M., Matthews, P. M., & McGonigle, D. J. (2005). Variability in fMRI: A re-examination of inter-session differences. *Hum. Brain Mapping*, *24*, 248–257.

Sokol, S. M., McCloskey, M., Cohen, N. J., & Aliminosa, D. (1991). Cognitive representations and processes in arithmetic: Inferences from the performance of braindamaged subjects. *J. Exp. Psychol. Learn. Mem. Cogn.*, *17*, 355–376.

Speer, N., Jacoby, L., & Braver, T. (2003). Strategy-dependent changes in memory: Effects on behavior and brain activity. *Cogn. Affective Behav. Neurosci.*, *3*, 155–167.

Squire, L. R. (1987). *Memory and brain.* New York: Oxford University Press.

Squire, L. R., Ojemann, J. G., Miezin, F. M., Petersen, S. E., Videen, T. O., & Raichle, M. E. (1992). Activations of the hippocampus in normal humans: A functional anatomical study of memory. *Proc. Natl. Acad. Sci. USA*, *89*, 1837–1841.

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.*, *27*, 279–306.

Stoff, D. M., & Eagle, M. N. (1971). The relationship among reported strategies, presentation rate, and verbal ability and their effects on free recall learning. *J. Exp. Psychol.*, *87*, 423–428.

Thorndike, E. L. (1911). *Individuality.* Boston: Houghton Mifflin.

Tsukiura, T., Mochizuki-Kawai, H., & Fujii, T. (2005). The effect of encoding strategies on medial temporal lobe activations during the recognition of words: An event-related fMRI study. *NeuroImage*, *25*, 452–461.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.

Tulving, E. (1989). Remembering and knowing the past. *Amer. Sci.*, *77*, 361–367.

Tulving, E., Habib, R., Nyberg, L., Lepage, M., & McIntosh, A. R. (1999). Positron emission tomography correlations in and beyond medial temporal lobes. *Hippocampus*, *9*, 71–82.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *J. Exp. Psychol.*, *70*, 122–129.

Vincent, J. L., Snyder, A. Z., Fox, M. D., Shannon, B. J., Andrews, J. R., Raichle, M. E., & Buckner, R. L. (2006). Coherent spontaneous activity identifies a hippocampal-parietal memory network. *J. Neurophysiol.*, *96*, 3517–3531.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*, 329–338.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*, 1188–1191.

Warnking, J., Dojat, M., Guérin-Dugué, A., Delon-Martin, C., Olympieff, S., Richard, N., Chéhikian, A., & Segebarth, C. (2002). fMRI retinotopic mapping—Step by step. *NeuroImage*, *17*, 1665–1683.

Weinstein, C. E., Underwood, V. L., Wicker, F. W., & Cubberly, W. E. (1979). Cognitive learning strategies: Verbal and imaginal elaboration. In H. F. O'Neil & C. D. Spielberger (Eds.), *Cognitive and affective learning strategies* (pp. 45–75). New York: Academic Press.

Wittenberg, G. M., & Tsien, J. Z. (2002). An emerging molecular and cellular framework for memory processing by the hippocampus. *Trends Neurosci.*, *25*(10), 501–505.

51 Constructive Memory and the Simulation of Future Events

DANIEL L. SCHACTER, DONNA ROSE ADDIS, AND RANDY L. BUCKNER

ABSTRACT Memory is widely conceived as a fundamentally constructive rather than purely reproductive process. One well-known source of evidence for constructive remembering is provided by various kinds of memory errors and illusions. A second line of evidence, which has recently emerged into the forefront of cognitive neuroscience, concerns the processes involved in imagining or simulating future events and novel scenes. In this chapter we discuss recent studies using various patient populations and neuroimaging techniques to examine future-event simulation and its relation to episodic memory, and we also link this research with earlier studies of constructive memory. Converging evidence supports the idea that imagining possible future events depends on much of the same neural machinery as does remembering past events, which we refer to as the core network. We consider conceptual and theoretical issues raised by this work, and also discuss adaptive functions of future-event simulation and related processes in the context of a constructive approach to memory.

> The first notion to get rid of is that memory is primarily or literally reduplicative, or reproductive. In a world of constantly changing environment, literal recall is extraordinarily unimportant . . . memory appears to be an affair of construction rather than reproduction.
>
> —Bartlett, 1932, pp. 204–205

When Sir Frederic Bartlett drew on experimental observations of errors and distortions in recall of complex stories to argue that memory is a fundamentally constructive process, his claims had little influence on his contemporaries. Psychological research on memory at the time was dominated by studies of rote learning in simple paired-associate paradigms; Bartlett's methods and theories made little sense in the context of the prevailing behaviorist zeitgeist. Several decades passed before Bartlett's ideas about constructive memory, revived by the publication of Neisser's (1967) seminal analysis of constructive processes in perception and memory, began to receive support from cognitive studies during the 1970s (for historical reviews, see Roediger, 1996; Schacter, 1995). Since that time, overwhelming cognitive evidence has accumulated in favor of Bartlett's claim that memory is "an affair of construction rather than reproduction" (for overviews, see Brainerd & Reyna, 2005; Loftus, 2003; Schacter, 1996, 2001).

When we turn to the cognitive neuroscience of memory, the situation looks a bit different. While cognitive neuroscientists have not opposed the idea that memory involves constructive processes, sustained interest in constructive aspects of memory has developed only recently. Of course, neurologists and neuropsychologists have long been interested in the phenomenon of confabulation, where patients with damage to various regions within prefrontal cortex and related regions produce vivid but highly inaccurate "recollections" of events that never happened. Clinicians have produced striking clinical reports of confabulation (e.g., Talland, 1961), and more recently a number of investigators have approached the phenomenon experimentally (for review, see Schnider, 2008). During the past decade, investigations of memory distortions in other patient populations, as well as neuroimaging studies of accurate versus inaccurate remembering in healthy individuals, have greatly increased our understanding of the cognitive neuroscience of constructive memory (Schacter, Norman, & Koutstaal, 1998; Schacter & Slotnick, 2004).

Even more recently—during just the past few years—there has been a dramatic increase in research on a related topic that also illuminates the constructive nature of memory: the role of memory in imagining or simulating possible future events (cf., Buckner & Carroll, 2007; Buckner, Andrews, & Schacter, 2008; Schacter, Addis, & Buckner, 2007, 2008; Suddendorf & Corballis, 2007). Evidence has rapidly accumulated to support the idea that memory—especially episodic memory, the system that allows individuals to recollect past events—is also critically involved in our ability to imagine future happenings and carry out related kinds of mental simulations. Furthermore, brain regions traditionally identified with memory, including the hippocampus, are similarly engaged when people carry out various

DANIEL L. SCHACTER Department of Psychology, Harvard University, Cambridge, Massachusetts; Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts

DONNA ROSE ADDIS Department of Psychology, University of Auckland, Auckland, New Zealand

RANDY L. BUCKNER Department of Psychology, Center for Brain Sciences, Harvard University, Cambridge, Massachusetts; Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts; Howard Hughes Medical Institute, Cambridge, Massachusetts

mental simulations. These investigations have provided new evidence concerning constructive processes in memory and may even provide clues concerning the functions of such constructive processes (Hassabis, Kumaran, & Maguire, 2007; Schacter & Addis, 2007a, 2007b).

One important impetus for this new wave of research emerged from claims made by Tulving (1983, 2002) that episodic memory supports "mental time travel" in both the past and the future. Tulving, influenced by prior related ideas from the Swedish neuroscientist David Ingvar (1979, 1985; for further discussion, see Buckner & Carroll, 2007; Schacter et al., 2008), suggested that mental time travel supports "autonoetic," or self-knowing, consciousness, which allows individuals to view themselves as temporally extended entities whose present awareness is influenced by the recollected past and imagined future.

During the 1990s, these ideas about mental time travel became associated with discussions concerning whether this capacity is unique to human beings or whether nonhuman animals also possess some form of autonoetic consciousness that allows them to revisit the past and anticipate the future. Suddendorf and Corballis (1997, 2007) and Tulving (2002, 2005) have both argued forcefully that mental time travel is restricted to human beings. While both Tulving (2002, 2005) and Suddendorf and Corballis (1997, 2007) allow that nonhuman animals can use semantic or procedural memory systems to gain access to stored information, they assert that such processes need not involve either recollecting a past event or using mental simulation to "preexperience" a future event—the essence of mental time travel.

This strong claim has spurred considerable debate (see Clayton, Bussey, & Dickinson, 2003, and commentaries on Suddendorf & Corballis, 2007) and will likely be difficult to resolve definitively owing to limitations on our ability to assess inner experience in nonhumans, which is central to the concept of mental time travel. Some compelling experimental demonstrations, at the very least, cast doubt on the strong claim for human uniqueness. For example, Clayton and Dickinson (1998) showed that food-caching scrub jays are able to retrieve detailed information about what food they cached as well as when and where they cached it, and Raby, Alexis, Dickinson, and Clayton (2007) have shown that conditions exist in which jays cache food in a way that appears to indicate some type of planning for the future. In a related line of research with rodents, several investigators have provided evidence indicating some of type of prospective coding, including evidence that hippocampal neurons encode not only a rat's current location and recent memory, but also encode prospective information concerning where the rat needs to go in the immediate future (Diba & Buzsáki, 2007; Ferbinteanu & Shapiro, 2003; Foster & Wilson, 2006; A. Johnson & Redish, 2007; Pastalkova, Itskov, Anavasingham, & Buzsáki, 2008). Whether or not such observations indicate the occurrence of mental time travel in rats, they do suggest that the hippocampus may provide prospective signals that could be used as a basis for making decisions.

Perhaps overlooked in the intensive discussion over whether animals can engage in mental time travel is that relatively little is known about how humans use memory to imagine or simulate future events. Although cognitive neuroscience has made much progress in delineating the nature of remembering, it has barely scratched the surface in studying how memory is used to imagine future events and to engage in related forms of mental simulation. The upsurge in relevant research during the past few years has begun to rectify the situation. In this chapter, we will focus on recent cognitive neuroscience research that has examined relations among memory, imagination, and future-event simulation.

Imagining future events: Findings and ideas

Insights into the nature of future-event simulation have been gained by cognitive studies of healthy young adults and memory-impaired populations, and more recently by neuroimaging studies. In the present chapter, we focus on memory-impaired populations and neuroimaging studies (for more general reviews, see Buckner et al., 2008; Schacter et al., 2008).

Studies of Future-Event Simulation in Memory-Impaired Populations

We consider here three memory-impaired populations in which future-event simulation has been examined: amnesic patients, older adults, and psychopathological populations.

Amnesic patients It is well-established that the amnesic syndrome resulting from damage to the medial temporal lobes and related structures is associated with a severe impairment in the ability to remember past experiences (see chapter 46 by Shrager and Squire, this volume). Early clinical observations (Talland, 1965) suggested that amnesic patients might also have problems envisioning their personal futures and planning for upcoming events. Tulving (1985) reported that the densely amnesic patient KC, who cannot remember any specific episodes from his past (for a review of KC, see Rosenbaum et al., 2005), exhibits similar problems envisioning any specific episodes in his future (Rosenbaum, Gilboa, Levine, Winocur, & Moscovitch, in press; Tulving, 1985; Tulving, Schacter, McLachlan, & Moscovitch, 1988). Note, however, that KC is characterized by fairly extensive brain damage, including damage to medial temporal, prefrontal, and other regions (see Rosenbaum et al.), thereby limiting the specificity with which his problems remembering the past or imagining the future can be associated with particular brain regions. A similar issue applies to a later and more

systematic study by Klein, Loftus, and Kihlstrom (2002) concerning patient DB, who became amnesic as a result of cardiac arrest and consequent anoxia. DB showed marked deficits on a 10-item questionnaire probing past and future events that were matched for temporal distance from the present (e.g., "What did you do yesterday? What are you going to do tomorrow?"). The patient's deficit in simulating future events appeared to involve only his personal future, since DB showed little difficulty imagining possible future happenings in the public domain, such as political events.

More recently, Hassabis, Kumaran, Vann, and Maguire (2007) examined the ability of five patients with documented bilateral hippocampal amnesia to imagine novel experiences, such as "Imagine you're lying on a white sandy beach in a beautiful tropical bay." The experimenters scored the constructions of patients and controls based on the content, spatial coherence, and subjective qualities of the imagined scenarios. Four of the five hippocampal patients produced constructions that were significantly reduced in richness and content compared with those of controls, especially for the measure of spatial coherence. The single patient who performed normally on the imaginary scene task was characterized by some residual hippocampal tissue. Because the lesions in the other cases appear to specifically include the hippocampal formation, this study strengthens the link between event simulation and hippocampal function. Note, however, that Hassabis and colleagues did not specifically require participants to imagine *future* events, indicating that the amnesic patients suffer from an impairment in event simulation that is not restricted to a particular time interval.

Older adults It is well known that healthy older adults exhibit a variety of episodic memory deficits (e.g., Craik & Salthouse, 2000), but little is known about future-event simulation in aging. Addis, Wong, and Schacter (2008) recently investigated the issue. They noted earlier work showing that aging is associated with reduced specificity during the recall of past autobiographical episodes. Levine, Svoboda, Hay, Winocur, and Moscovitch (2002) reported such age-related changes in the episodic quality of past events using the Autobiographical Interview (AI), a measure that distinguishes episodic information from other "external" details (e.g., semantic information, other external events, repetitions) that comprise a participant's description of a past event. Levine and colleagues observed that older adults recalled significantly fewer internal/episodic details and tended to produce more external/semantic information.

Addis, Wong, and Schacter (2008) used an adapted version of the Autobiographical Interview that required young and older participants to generate memories of past events and simulations of future events in response to individual word cues. They allowed participants three minutes to describe each episode, and transcriptions of the events were segmented into distinct details that were classified as either internal (episodic) or external (semantic). The key finding was that older adults generated fewer internal details than younger adults; importantly, this effect was observed to the same extent for future events as for past events. By contrast, older adults showed small but significant increases relative to young adults in the production of external details for both past and future events. Furthermore, there were strong positive correlations across past and future events for both internal and external detail scores, whereas internal and external detail scores were not correlated with one another. Finally, the internal (but not external) detail score correlated significantly with a measure of relational memory (paired-associate learning), known to be dependent on the hippocampus, a point to which we will return later when discussing theoretical accounts of future-event simulation. Overall, the results reveal a strong link between remembering the past and imagining the future in older adults. These findings dovetail nicely with observations from Spreng and Levine (2006), who reported similar temporal distributions for past and future events in aging: when remembering past events or imagining future events that are likely to happen, both older and younger adults generated the highest number of events near the present, with the frequency declining as a function of time in a manner well described by a power function.

Psychopathological populations A growing number of studies have examined future-event simulation in patients with various forms of psychopathology. We have reviewed this literature in detail elsewhere (Schacter et al., 2008) and summarize several key findings here. Williams and colleagues (1996) reported a seminal study in which they found that suicidally depressed patients have difficulty recalling specific memories of past events and also in generating specific simulations of future events. Compared to nondepressed controls, the past and future events generated by depressed patients in response to cue words lacked specific detail and thus were characterized as "overgeneral"; these reductions in specificity of past and future events were significantly correlated. Similar findings have been reported in milder forms of depression (e.g., Dickson & Bates, 2005; MacLeod, Rose, & Williams, 1993) and also in anxious individuals (Stöber & Borkovec, 2002).

Williams and colleagues (1996) found that past and future events generated by suicidally depressed patients were overgeneral for both positive and negative events. However, others have reported effects of valence. For instance, MacLeod and colleagues (1993) found that suicidally depressed patients were less able to envision *positive* future episodes (see also Dickson & Bates, 2006). Indeed, reduced access to positive future events correlates with the severity of hopelessness (MacLeod & Cropley, 1995), suggesting that

simulation deficits may help to maintain the sense of hopelessness that typically characterizes depression (for related neuroimaging research concerning neural correlates of optimism, see Sharot, Riccardi, Raio, & Phelps, 2007, and commentary by Schacter & Addis, 2007c). Similarly, increased access to simulations of negative future events is characteristic of anxiety disorders (e.g., MacLeod, Tata, Kentish, Carroll, & Hunter, 1997; Ruane, MacLeod, & Holmes, 2005).

Such observations in patients with depression and anxiety disorders have led to the proposal that the reduced specificity of autobiographical memories and future-event simulations reflects problems with affect regulation: patients produce overgeneral events because they truncate search or construction to protect themselves from experiencing potentially destabilizing memories or simulations (Williams, 1996, 2006). Recently, D'Argembeau, Raffard, and Van der Linden (2008) reported that schizophrenics generated significantly fewer specific past and future events than did healthy controls. Such findings are less likely to be attributable to the kinds of affect-regulation problems that occur in depression and anxiety. Moreover, the findings from depression, anxiety, and schizophrenia are quite similar to those considered earlier from amnesic patients and older adults.

Taken together, these observations encourage further consideration of the possible role of neuropsychological deficits that may contribute to the reduced specificity of events evident in both psychiatric and nonpsychiatric populations. For example, we have suggested (Schacter et al., 2008) that the aforementioned data from amnesic patients (and data to be considered shortly from neuroimaging studies) implicating the hippocampus in event simulation raise the possibility that hippocampal dysfunction might contribute to overgeneral simulations of past and future events. Hippocampal atrophy is evident in a number of psychiatric conditions in which simulation deficits have been documented, including depression (Bremner et al., 2000; Campbell & Macqueen, 2004) and schizophrenia (Velakoulis et al., 2006), and it also has been documented in older adults (e.g., Driscoll et al., 2003; Golomb et al., 1993). It is therefore possible that hippocampal dysfunction contributes to simulation deficits observed across these varied populations.

Neuroimaging of Future-Event Simulation During the past couple of years, several studies have used neuroimaging techniques to compare the neural correlates of imagining future events with those that characterize remembering past events. We first review key experimental findings and related observations before turning to some emerging conceptual issues.

Basic findings: The core network A consistent finding across studies has been that remembering the past and imagining the future recruit a similar network of brain regions. Such findings were reported initially in an early positron emission tomography (PET) study from Okuda and colleagues (2003; see also Partiot, Grafman, Sadato, Wachs, & Hallett, 1995, for related early findings). During scanning, participants talked freely about either the near past or future (i.e., the last or next few days) or the distant past or future (i.e., the last or next few years). Similar levels of activation were observed during past and future conditions in several prefrontal regions, as well as in the medial temporal lobe (right hippocampus and bilateral parahippocampal gyrus). Note, however, that because Okuda and colleagues used a relatively unconstrained paradigm that did not probe participants about particular events, it is unclear whether these reports consisted of episodic memories and simulations (unique events specific in time and place) or general semantic information about an individual's past or future. More recent fMRI studies have used event-related designs to yield information regarding the neural bases of specific past and future events.

Szpunar, Watson, and McDermott (2007) instructed participants to remember specific events that occurred in their personal past, imagine specific future events that might occur in their personal future, or imagine specific events involving a familiar individual (Bill Clinton) in response to event cues (e.g., past birthday, retirement party). Consistent with previous observations, there was considerable overlap in activity associated with past and future events in the bilateral frontopolar and medial temporal lobe regions, as well as in posterior cingulate cortex. Note also that these regions were not activated to the same degree when participants imagined events involving Bill Clinton, seeming to demonstrate a neural signature that is unique to the construction of events in one's *personal* past or future.

One general issue that applies to the foregoing studies, and potentially to any neuroimaging study that compares the neural correlates of remembering past events and imagining future events, is that remembering is usually associated with greater levels of episodic detail than is imagining (e.g., M. Johnson, Foley, Suengas, & Raye, 1988). To the extent that this outcome occurs, comparisons between past and future events will be partly or entirely confounded by differences in level of detail. Using event-related fMRI, Addis, Wong, and Schacter (2007) attempted to equate experimentally the level of detail and related phenomenological features of past and future events. Also, taking advantage of the temporal resolution of fMRI, the past and future tasks were divided into two phases: (1) an initial construction phase during which participants generated a past or future event in response to an event cue (e.g., "dress") and pressed a button when they had an event in mind, and (2) an elaboration phase during which participants generated as much detail as they could about the event.

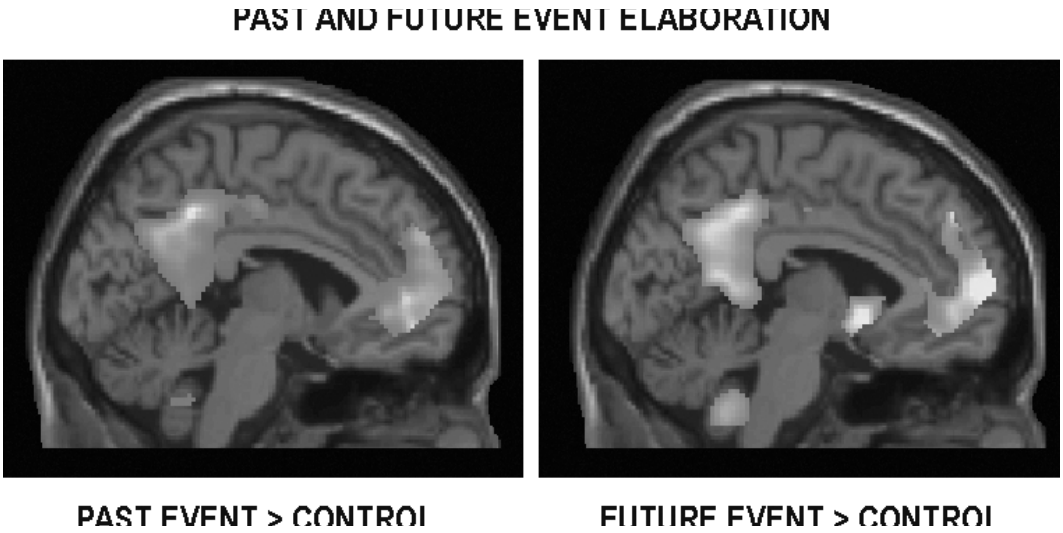


FIGURE 51.1 Sagittal slice ($x = -4$) illustrating the striking commonalities in the medial left prefrontal and parietal regions engaged when remembering the past (*left panel*) and imagining the future (*right panel*). These marked similarities of activation were also evident in areas of the medial temporal lobe (left hippocampus, bilateral parahippocampal gyrus) and lateral cortex (left temporal pole and left bilateral inferior parietal cortex). This extensive pattern of common activity was not present during the construction of past and future events; it only emerged during the elaboration of these events (shown here, relative to the elaboration phase of a semantic and an imagery control task; significant at $p < .001$, uncorrected; shown at $p < .005$, uncorrected.) (Originally published in Addis, Wong, & Schacter, 2007.) (See color plate 63.)

The construction phase was associated with some common past-future activity in posterior visual regions and left hippocampus. During the elaboration phase, when participants focused on generating details about the remembered or imagined event, there was even more extensive overlap between the past and future tasks (see figure 51.1). Both event types were associated with activity in a network of regions including medial temporal (hippocampus and parahippocampal gyrus) and prefrontal cortex, as well as medial parietal and retrosplenial cortex.

Botzung, Denkova, and Manning (2008) have recently reported data from an fMRI study that are mainly consistent with those from the preceding studies. The day before scanning, subjects initially reported on 20 past events from the last week and 20 future events planned for the next week. The experimenters constructed cue words for these events that were presented to subjects the next day during scanning, when they were instructed to think of past or future events to each cue. Past and future events produced activation in a network similar to that reported by Addis and colleagues (2007).

Collectively, the results from the preceding imaging studies consistently implicate a core network of structures in both remembering the past and imagining the future (Buckner & Carroll, 2007; Buckner et al., 2008; Schacter et al., 2007, 2008). This network consists of prefrontal and medial temporal lobe regions, as well as posterior regions including lateral parietal, posterior cingulate, and retrosplenial cortices that have previously been observed as components of brain networks important to memory retrieval (Cabeza & St Jacques, 2007; Gilboa, 2004; Maguire, 2001; Spreng, Mar, & Kim, in press; Svoboda, McKinnon, & Levine, 2006; Wagner, Shannon, Kahn, & Buckner, 2005). Analyses of the interactions among the brain regions within this core network demonstrate that all of the component regions are selectively correlated with one another within a large-scale brain system that includes the hippocampal formation (Greicius, Srivastava, Reiss, & Menon, 2004; Kahn, Andrews-Hanna, Vincent, Snyder, & Buckner, 2008; Vincent et al., 2006), and that the network likely consists of distinct interacting subsystems (Buckner et al., 2008).

Although it seems clear that remembering the past and imagining the future are both associated to some extent with a common core network, neuroimaging studies have also yielded a number of findings that point to possible differences between the two. First, direct comparisons have consistently shown greater activity in several brain regions when individuals imagine the future than when they remember the past. For example, Okuda and colleagues (2003) reported greater activity in frontopolar and medial temporal regions when people talked about the future than the past; Szpunar and colleagues (2007) reported that bilateral premotor cortex and left precuneus were more active for future relative to past events, but not vice versa; and Addis and colleagues (2007) found that during the early construction phase of future simulation, several regions showed greater activity for future versus past events (but not the reverse), including right hippocampus and frontopolar cortex.

In a more recent study, Addis, Cheng, and Schacter (2008) contrasted activity when individuals were cued to remember or imagine specific events, as in previous studies, versus when they were cued to remember general, routine events (e.g., having brunch after attending church) or to imagine generic events that might occur sometime in their personal futures (e.g., reading the newspaper each morning). Addis and colleagues replicated the foregoing findings of greater activity for future than past events during the early phase of event construction. Furthermore, they found that the left frontal pole showed this future > past effect for both specific and generic events, suggesting a general role in prospection irrespective of the specificity of the event. However, the right hippocampus showed the future > past effect only for specific events; in fact, there was no evidence for right hippocampal activity during construction of generic future events.

These observations are open to multiple interpretations (note also that Botzung et al., 2008, reported evidence for increased activity for past versus future events; but see Schacter et al., 2008, for discussion of methodological issues that complicate intepretation of this finding). For example, Szpunar and colleagues (2007) suggested that a more active type of imagery processing might be required by future than past events. Addis and colleagues (2007) hypothesized that more intensive constructive processes are required by imagining future events than by retrieving past events. While both past- and future-event tasks require the retrieval of information from memory, thus engaging common memory networks, only the future task requires that event details gleaned from various past events be flexibly recombined into an imaginary event, perhaps resulting in increased activity during future-event tasks. A related possibility is that imagined future events are more novel than remembered past events; increased activity during future-event tasks might reflect some form of novelty encoding. This latter idea is potentially applicable to findings of increased hippocampal activation during future-event tasks, since it is well known that encoding novel events can be associated with increased hippocampal activity (e.g., Ranganath & Rainer, 2003). Note, however, that Addis, Cheng, and Schacter's (2008) finding that increased right hippocampal activity for future events was observed for specific but not generic events would appear to be inconsistent with a simple novelty-encoding account, because both the specific and generic future events were novel.

Addis and Schacter (2008) report additional findings concerning differential neural responses to past and future events. They conducted parametric modulation analyses, with temporal distance and detail as covariates, focusing on the hippocampal and the frontopolar regions. They hypothesized that *reintegrating* increasing amounts of detail for either a past or future event would be associated with increasing levels of hippocampal activity. By contrast, because future events are thought to require more intensive *recombining* of disparate details into a coherent event, the hippocampal response to increasing amounts of future-event detail should be larger than that for past-event detail. In addition, since the frontal pole is thought to play a role in prospective thinking (e.g., Okuda et al., 2003), this region should also exhibit a future > past detail response if it is specifically involved in the generation of future details.

Consistent with predictions, the analysis showed that the left posterior hippocampus was responsive to the amount of detail comprising both past and future events. In contrast, a separate region in the left anterior hippocampus responded differentially to the amount of detail comprising future events, possibly reflecting the recombination of details into a novel future event. Moreover, the right frontal pole responded significantly more to the generation of future- relative to past-event details, again suggesting that this region might be involved specifically in prospective thinking.

The parametric modulation analysis of temporal distance revealed that the increasing recency of past events was associated with activity in the right parahippocampus gyrus (BA 35/36), while activity in the bilateral hippocampus was associated with the increasing remoteness of future events. Addis and Schacter (2008) proposed that the hippocampal response to the distance of future events reflects the increasing disparateness of details likely included in remote future events and the intensive relational processing required for integrating such details into a coherent episodic simulation of the future. More generally, these results suggest that the core network supporting past- and future-event simulation can be recruited in different ways depending on whether the generated event is in the past or future.

Conceptual issues: Past versus future or remembering versus imagining? The preceding observations raise a general point concerning the growing number of studies that have compared remembering the past with imagining the future. When differences between these two conditions are observed, they are typically attributed to differences in the way that the brain handles past and future events. However, in the reviewed studies past events are remembered whereas future events are imagined; accordingly, the differences could equally well be attributed to differences between remembering and imagining, rather than differences between past and future per se. Of course, the future cannot be remembered because it has not yet happened. However, both the past and the future can be imagined. Furthermore, events can be imagined without any specific reference to a particular time point. Therefore, it would be of interest to determine whether any of the foregoing findings are indeed specifically related to imagining future events, or whether such findings are observed when people imagine events (1) that lack a specific

temporal reference or (2) that might have occurred in their personal pasts. Recent studies provide evidence concerning both points.

Hassabis, Kumaran, and Maguire (2007) adapted the experimental paradigm that they had used previously with amnesic patients for an fMRI study with healthy volunteers in which participants were asked to imagine novel, fictitious scenes, without explicit reference to whether those scenes should be placed in the past, present, or future. Subjects were then scanned in a subsequent session in which they were cued to remember the previously constructed fictitious scenes, construct additional novel fictitious scenes, or recall real episodic memories from their personal pasts. Hassabis and colleagues found that all three conditions were associated with activations in some of the regions within the core network that were associated with future-event simulation in previously reviewed studies, including hippocampus, parahippocampal gyrus, and retrosplenial cortex. The results thus indicate that activity in these regions is not restricted to conditions that explicitly require imagining future events. However, Hassabis and colleagues also reported that remembering "real" episodic memories yielded increased activity in several core network regions—notably anterior medial prefrontal cortex and posterior cingulate cortex—in comparison with constructing fictitious events. We will return shortly to the theoretical implications of these latter findings.

In a related study, Addis, Pan, Vu, Laiser, and Schacter (in press) attempted to disambiguate whether future-event-related activity is specifically associated with prospective thinking or with the more general demands of imagining an episodic event in either temporal direction by instructing subjects to imagine events that might occur in their personal future or events that might have occurred in their personal pasts. Prior to scanning, participants provided episodic memories of actual experiences that included details about a *person, object, and place* involved in that event. During scanning, the subjects were cued to recall some of the events that had actually occurred, and for the conditions in which they imagined events, the experimenters randomly recombined details concerning person, object, and place from separate episodes. Participants were thus presented with cues for a person, object, and place taken from multiple episodes, and were instructed to imagine them together in a single, novel episode that included the specified details.

Addis and colleagues (in press) reported that all regions within the core network (including medial prefrontal and frontopolar cortex, hippocampus, parahippocampal gyrus, lateral temporal and temporopolar cortex, medial parietal cortex including posterior cingulate and restrosplenial cortex, and lateral parietal cortex) were similarly engaged when participants imagined future and past events, suggesting that the network can be used for event simulation regardless of the temporal location of the event.

Theoretical implications: Future event simulation and constructive memory

Neuroimaging and neuropsychological observations have led to a number of new theoretical proposals, involving both attempts to describe the critical cognitive processes associated with core network activation and attempts to consider functional aspects of future-event simulation. We have reviewed these proposals in detail elsewhere (Buckner & Carroll, 2007; Buckner et al., 2008; Schacter et al., 2007, 2008). Here, we briefly summarize the main ideas. We first consider two related attempts to delineate key processes associated with the core network, and then describe a related idea that attempts to link the core network and future-event simulation with memory errors and related constructive aspects of remembering.

Core Network Activation: Critical Cognitive Processes The experimental work that we have reviewed has sparked a number of attempts to characterize the key processes subserved by the core network that is consistently activated in recent studies of imagining and remembering (Buckner & Carroll, 2007; Hassabis, Kumaran, & Maguire, 2007; Hassabis, Kumaran, Vann, et al., 2007; Hassabis & Maguire, 2007; Schacter & Addis, 2007a, 2007b; Spreng et al., in press). As we have noted in our previous reviews, these perspectives share much in common and differ mainly in points of emphasis and focus.

Buckner and Carroll (2007) and Buckner and colleagues (2008) argued that the core network serves a common set of processes by which past experiences are used adaptively to imagine perspectives and events beyond those that emerge from the immediate environment. By this view, the functions of the core network are not restricted to tasks requiring mental time travel. In addition to the network's role in remembering the past and envisioning the future, the core network is hypothesized to contribute to more general functions, extending to diverse tasks that require mental simulation of alternative perspectives. They observed that some, but not all, regions within the core network are engaged during theory-of-mind tasks that require thinking about the perspectives of others (e.g., Saxe & Kanwisher, 2003), and they also noted that such regions may be engaged during certain kinds of spatial navigation tasks (e.g., Byrne, Becker, & Burgess, 2007). Buckner and Carroll suggested that the core brain network is commonly engaged when individuals are simulating alternative perspectives, including alternatives in the present and possibilities in the future—a process they provisionally termed *self-projection*. This view predicts that activation of the core network should correspond to the

extent that a task encourages simulation of an alternative perspective not beyond the immediate environment. Spreng and colleagues (in press) performed a meta-analysis of studies that generally supported this broad hypothesis.

In a detailed analysis of the anatomy and functional connections among the regions within the network, Buckner and colleagues (2008) recently expanded on this perspective and showed that the core network comprises *at least* two interacting subsystems: the medial temporal lobe subsystem functions to provide information from memory; the dorsal medial prefrontal cortex subsystem participates to derive self-relevant mental simulations. The two subsystems interact through hubs, including the posterior cingulate, but can be dissociated using the observation that dorsal medial prefrontal cortex and the medial temporal lobe are not intrinsically correlated with one another. One possibility is that mental simulations such as remembering and envisioning the future draw heavily on contributions from both subsystems, whereas other forms of task rely preferentially on one subsystem. For example, theory-of-mind tasks that do not draw on memory rely primarily on the medial prefrontal subsystem, as evidenced by strong activation of that system and not the medial temporal lobe. Consistent with this idea, patients with medial temporal lesions exhibit intact performance on theory-of-mind tasks that do not draw on past memories (Rosenbaum, Stuss, Levine, & Tulving, 2007). However, the neuropsychological data also indicate that patients with damage to medial prefrontal regions (Bird, Castelli, Malik, Frith, & Husain, 2004) show intact performance on several theory-of-mind tasks, which is perplexing in light of the common activation of this region in imaging studies.

Hassabis and Maguire (2007) have argued that a process they refer to as *scene construction* links together various tasks that depend on many regions within the core network, in particular those associated with the medial temporal subsystem. Scene construction focuses on visuospatial aspects of mental simulations and was motivated initially by the previously discussed finding that amnesic patients with medial temporal damage show deficits when asked to imagine novel scenes, with a disproportionate impairment in the spatial coherence of the imagined scenes (Hassabis, Kumaran, Vann, et al., 2007). Neuroimaging findings from the same task likewise show core network activity (Hassabis, Kumaran, & Maguire, 2007). Because the novel-scenes task does not explicitly require mental time travel, Hassabis and Maguire contended that projecting oneself into the past or the future is not the critical process for activating the medial temporal subsystem.

Taken in the context of the anatomic analysis of Buckner and colleagues (2008), the collective results begin to converge on the idea that the core network comprises at least two subsystems that interact to accomplish autobiographical remembering and envisioning the future but that are also used to varying degrees across a diverse set of tasks that extend well beyond forms of "mental time travel" (for further discussion of the neural correlates of mental time travel, see Arzy, Molnar-Szarkacs, & Blanke, 2008).

Episodic Simulation, the Core Network, and Constructive Memory Schacter and Addis (2007a, 2007b; for related ideas, see Dudai & Carruthers, 2005; Suddendorf & Corballis, 1997; Suddendorf & Busby, 2005) have linked findings concerning event simulation and core network activity to the observation that memory involves a constructive process of piecing together bits and pieces of information. According to the *constructive episodic simulation hypothesis*, imagining future events requires a system that can flexibly recombine details from past events. From this perspective, past and future events draw on similar information stored in episodic memory and rely on similar underlying processes; episodic memory supports the construction of future events by extracting and recombining stored information into a simulation of a novel event. The adaptive value of such a system is that it enables past information to be used flexibly in simulating alternative future scenarios without engaging in actual behavior. A potential downside of such a system, however, is that it is vulnerable to memory errors, such as misattribution and false recognition (for examples, see Schacter & Addis, 2007a, 2007b). This observation suggests, intriguingly, that certain kinds of memory errors may be the by-product of a system whose adaptive function is to make available information from the past in a flexible form that supports simulations of future events.

The constructive episodic simulation hypothesis receives general support from the previously reviewed findings of neural and cognitive overlap between past and future events; and, because it emphasizes the importance of flexibly relating and recombining information from past episodes, the hypothesis is more specifically supported by the mounting evidence from amnesia (Hassabis, Kumaran, Vann, et al., 2007), neuroimaging (Addis et al., 2007; Addis & Schacter, 2008; Botzung et al., 2008; Hassabis, Kumaran, & Maguire, 2007; Okuda et al., 2003), and aging (Addis, Wong, & Schacter, 2008), linking hippocampal function and relational processing with episodic simulation. The hippocampal region is thought to support relational memory processes (e.g., Eichenbaum & Cohen, 2001); and, according to the constructive episodic simulation hypothesis, these processes are critical for recombining stored information into future-event simulations.

Because the constructive episodic simulation hypothesis places great emphasis on the process of recombining event details, it is critical to determine whether such recombination

processes are critical for future-event simulation, or whether such simulations are based on retrieval of entire past episodes, or fragments of such episodes, which are simply recast as possible future events. Relevant data are provided by the aforementioned study by Addis and colleagues (in press) using experimental recombination of details from distinct episodes: core network activation, including the hippocampus, was observed under conditions that effectively ruled out recasting of a single past episode as a future event.

Although further research is required to delineate the exact role of the hippocampus in mental simulation, it may be worth noting that research on other aspects of constructive memory has also highlighted the involvement of the hippocampus and related medial temporal lobe regions. For instance, some neuroimaging studies of false recognition, where individuals claim to have previously encountered a novel item that is conceptually or perceptually related to a previously studied item, have documented hippocampal/medial temporal lobe activation during false recognition of semantically associated words (e.g., Cabeza, Rao, Wagner, Mayer, & Schacter, 2001) or abstract shapes (Slotnick & Schacter, 2004). Similarly, several studies have shown that amnesic patients with medial temporal lobe damage show reduced levels of false recognition for various kinds of materials, suggesting that the hippocampal region is involved with encoding and/or retrieving the information that drives false recognition effects (e.g., Schacter, Verfaellie, & Pradere, 1996; Verfaellie, Page, Orlando, & Schacter, 2005). Taken together with the evidence for hippocampal involvement during simulation of future or novel events, it seems increasingly clear that the hippocampus is related importantly to constructive aspects of memory.

Studies of future-event simulation also bring into sharp focus fundamental issues concerning processes of *reality monitoring*, which allow us to distinguish between remembered and imagined events (Johnson & Raye, 1981). If remembering past events and imagining future or novel events recruit largely overlapping brain networks, how can individuals distinguish fantasy from reality?

Hassabis, Kumaran, and Maguire (2007) addressed this issue in the context of their neuroimaging study, where they found, as noted earlier, that anterior medial prefrontal cortex and posterior cingulate cortex showed greater activity when individuals recollected real episodic memories as compared to when they imagined novel scenes. Because their novel-scenes task does not require mental time travel or projection of the self, Hassabis, Kumaran, and Maguire suggested that anterior medial prefrontal cortex and posterior cingulate cortex "support episodic memory over and above scene construction" (2007, p. 14372), perhaps contributing to effective reality monitoring. While this conclusion may be accurate in the context of the scene-construction task, these regions typically activate as part of the core network when individuals imagine themselves in personal future events (Addis et al., 2007; Szpunar et al., 2007). Interestingly, in a recent study (Abraham, von Cramon, & Schubotz, 2008) where participants were asked to imagine scenarios that involved meeting real people (e.g., George Bush) versus fictional characters (e.g., Cinderella), anterior prefrontal and posterior cingulate were more active during the former than the latter condition, possibly indicating greater ease of self-projection when imagining oneself meeting an actual person (medial temporal regions were similarly active in the two conditions). Taken together, the foregoing studies suggest that additional areas and processes (beyond anterior medial prefrontal and posterior cingulate) must be recruited to allow one to distinguish an episodic memory from a realistic future simulation that engages the self.

Here, it seems likely that there is a role for the long-standing idea from research on reality monitoring that remembering events that one has actually experienced is associated with greater numbers of sensory and perceptual details than remembering previously imagined events (e.g., Johnson & Raye, 1981). This idea has received support from behavioral studies (e.g., Johnson et al., 1988) as well as neuromaging research (Kensinger & Schacter, 2006; see also Slotnick & Schacter, 2004). Most directly related to the present concerns, Addis and colleagues (in press) report preliminary evidence that remembering actual autobiographical events is more strongly associated with activity in posterior visual cortex (and some medial temporal regions) than is imagining future or past events using the previously described procedure of cuing imagined events by recombining details from different actual events. In this study, remembered events were rated as significantly more detailed than imagined events, so it would make sense from the perspective of the reality-monitoring framework that regions associated with processing of sensory and contextual details would show greater activity for real events than for imagined ones.

Although still in its infancy, it seems clear that research on future-event simulation and related forms of internally directed cognition has much to offer memory research. At the very least, the striking similarities observed during remembering the past and imagining the future are consistent with Bartlett's (1932) claim that "memory appears to be an affair of construction rather than reproduction." We are optimistic that further study of such processes as future-event simulation, scene construction, and self-projection will teach us much about the constructive nature of memory.

ACKNOWLEDGMENTS Preparation of this chapter was supported by grants from the NIA, NIMH, and HHMI. We thank Adrian Gilmore for help with preparation of the manuscript.

REFERENCES

Abraham, A., von Cramon, D. Y., & Schubotz, R. I. (2008). Meeting George Bush versus meeting Cinderella: The neural response when telling apart what is real from what is fictional in the context of our reality. *J. Cogn. Neurosci., 20*, 965–976.

Addis, D. R., Cheng, T., & Schacter, D. L. (2008). *Episodic s imulation of specific and general future events*. Poster presented at the Annual Meeting of the Organization for Human Brain Mapping, Melbourne, Australia.

Addis, D. R., Pan, L., Vu, M. A., Laiser, N., & Schacter, D. L. (in press). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia*.

Addis, D. R., & Schacter, D. L. (2008). Constructive episodic simulation: Temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus, 18*, 227–237.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia, 45*, 1363–1377.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychol. Sci., 19*, 33–41.

Arzy, S., Molnar-Szarkacs, I. M., & Blanke, O. (2008). Self in time: Imagined self-location influences neural activity related to mental time travel. *J. Neurosci., 28*, 6502–6507.

Bartlett, F. C. (1932). *Remembering*. Cambridge, UK: Cambridge University Press.

Bird, C. M., Castelli, F., Malik, O., Frith, U., & Husain, M. (2004). The impact of extensive medial frontal lobe damage on theory of mind and cognition. *Brain, 127*, 914–928.

Botzung, A., Denkova, E., & Manning, L. (2008). Experiencing past and future personal events: Functional neuroimaging evidence on the neural bases of mental time travel. *Brain Cogn., 66*, 202–212.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York: Oxford University Press.

Bremner, J. D., Narayan, M., Anderson, E. R., Staib, L. H., Miller, H. L., & Charney, D. S. (2000). Hippocampal volume reduction in major depression. *Am. J. Psychiatry, 157*, 115–117.

Buckner, R. L., Andrews, J. R., & Schacter, D. L. (2008). The brain's default system: Anatomy, function, and relevance to disease. *The Year in Cognitive Neuroscience, Ann. NY Acad. Sci., 1124*, 1–38.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn. Sci., 11*, 49–57.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychol. Rev., 114*, 340–375.

Cabeza, R., Rao, S., Wagner, A. D., Mayer, A., & Schacter, D. L. (2001). Can medial temporal lobe regions distinguish true from false? An event-related fMRI study of veridical and illusory recognition memory. *Proc. Natl. Acad. Sci. USA, 98*, 4805–4810.

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends Cogn. Sci., 11*, 219–227.

Campbell, S., & Macqueen, G. (2004). The role of the hippocampus in the pathophysiology of major depression. *J. Psychiatry Neurosci., 29*, 417–426.

Clayton, N. S., Bussey, T. J., & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat. Rev. Neurosci., 4*, 685–691.

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature, 395*, 272–274.

Craik, F. I. M., & Salthouse, T. A. (Eds.). (2000). *Handbook of aging and cognition* (2nd ed.). Hillsdale, NJ: Erlbaum.

D'Argembeau, A., Raffard, S., & Van der Linden, M. (2008). Remembering the past and imagining the future in schizophrenia. *J. Abnorm. Psychol., 117*, 247–251.

Diba, K., & Buzsáki, G. (2007) Forward and reverse hippocampal place-cell sequences during replay. *Nat. Neurosci., 10*, 1241–1242.

Dickson, J. M., & Bates, G. W. (2005). Influence of repression on autobiographical memories and expectations of the future. *Aust. J. Psychol., 57*, 20–27.

Dickson, J. M., & Bates, G. W. (2006). Autobiographical memories and views of the future: In relation to dysphoria. *Int. J. Psychol., 41*, 107–116.

Driscoll, I., Hamilton, D. A., Petropoulos, H., Yeo, R. A., Brooks, W. M., Baumgarter, R. N., et al. (2003). The aging hippocampus: Cognitive, biochemical, and structural findings. *Cereb. Cortex, 13*, 1344–1351.

Dudai, Y., & Carruthers, M. (2005). The Janus face of Mnemosyne. *Nature, 434*, 823–824.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. New York: Oxford University Press.

Ferbinteanu, J., & Shapiro, M. L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron, 40*, 1227–1239.

Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioral sequences in hippocampal place cells during the awake state. *Nature, 440*, 680–683.

Gilboa, A. (2004). Autobiographical and episodic memory—One and the same? Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia, 42*, 1336–1349.

Golomb, J., de Leon, M. J., Kluger, A., George, A. E., Tarshish, C., & Ferris, S. H. (1993). Hippocampal atrophy in normal aging: An association with recent memory impairment. *Arch. Neurol., 50*, 967–973.

Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *Proc. Natl. Acad. Sci. USA, 101*, 4637–4642.

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *J. Neurosci., 27*, 14365–14374.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. USA, 104*, 1726–1731.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends Cogn. Sci., 11*, 299–306.

Ingvar, D. H. (1979). Hyperfrontal distribution of the cerebral grey matter flow in resting wakefulness: On the functional anatomy of the conscious state. *Acta Neurol. Scand., 60*, 12–25.

Ingvar, D. H. (1985). "Memory of the future": An essay on the temporal organization of conscious awareness. *Hum. Neurobiol., 4*, 127–136.

Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci., 27*, 12176–12189.

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *J. Exp. Psychol. Gen., 117*, 371–376.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychol. Rev., 88*, 67–85.

Kahn, I., Andrews-Hanna, J. R., Vincent, J. L., Snyder, A. Z., & Buckner, R. L. (2008). Distinct cortical anatomy linked to subregions of the medial temporal lobe revealed by intrinsic functional connectivity. *J. Neurophysiol, 100*, 129–139.

Kensinger, E. A., & Schacter, D. L. (2006). Neural processes underlying memory attribution on a reality-monitoring task. *Cereb. Cortex, 16*, 1126–1133.

Klein, S. B., Loftus, J., & Kihlstrom, J. F. (2002). Memory and temporal experience: The effects of episodic memory loss in an amnesic patient's ability to remember the past and imagine the future. *Soc. Cogn., 20*, 353–379.

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychol. Aging, 17*, 677–689.

Loftus, E. F. (2003). Make-believe memories. *Am. Psychol., 58*, 867–873.

MacLeod, A. K., & Cropley, M. L. (1995). Depressive future-thinking: The role of valence and specificity. *Cogn. Ther. Res., 19*, 35–50.

MacLeod, A. K., Rose, G., & Williams, J. M. (1993). Components of hopelessness about the future in parasuicide. *Cogn. Ther. Res., 17*, 441–455.

MacLeod, A. K., Tata, P., Kentish, J., Carroll, F., & Hunter, E. (1997). Anxiety, depression, and explanation-based pessimism for future positive and negative events. *Clin. Psychol. Psychother., 4*, 15–24.

Maguire, E. A. (2001). Neuroimaging studies of autobiographical event memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 356*, 1441–1451.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., et al. (2003). Thinking of the future and the past: The roles of the frontal pole and the medial temporal lobes. *NeuroImage, 19*, 1369–1380.

Partiot, A., Grafman, J., Sadato, N., Wachs, J., & Hallett, M. (1995). Brain activation during the generation of emotional and non-emotional plans. *NeuroReport, 6*, 1269–1272.

Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science, 321*, 1322–1327.

Raby, C. R., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). Planning for the future by western scrub-jays. *Nature, 445*, 919–921.

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci., 4*, 193–202.

Roediger, H. L., III. (1996). Memory illusions. *J. Mem. Lang., 35*, 76–100.

Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., & Moscovitch, M. (in press). Amnesia as an impairment of detail generation and binding: Evidence from personal, fictional, and semantic narratives in K.C. *Neuropsychologia*.

Rosenbaum, R. S., Kohler, S., Schacter, D. L., Moscovitch, M., Westmacott, R., Black, S. E., et al. (2005). The case of K.C.: Contributions of a memory-impaired person to memory theory. *Neuropsychologia, 43*, 989–1021.

Rosenbaum, R. S., Stuss, D. T., Levine, B., & Tulving, E. (2007). Theory of mind is independent of episodic memory. *Science, 318*, 1257.

Ruane, D., MacLeod, A. K., & Holmes, E. A. (2005). The simulation heuristic and visual imagery in pessimisim for negative events in anxiety. *Clin. Psychol. Psychother., 12*, 313–325.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NeuroImage, 19*, 1835–1842.

Schacter, D. L. (1995). Memory distortion: History and current status. In D. L. Schacter (Ed.), *Memory distortion: How minds, brains and societies reconstruct the past* (pp. 1–43). Cambridge, MA: Harvard University Press.

Schacter, D. L. (1996). *Searching for memory: The brain, the mind, and the past*. New York: Basic Books.

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. Boston: Houghton Mifflin.

Schacter, D. L., & Addis, D. R. (2007a). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 362*, 773–786.

Schacter, D. L., & Addis, D. R. (2007b). The ghosts of past and future. *Nature, 445*, 27.

Schacter, D. L., & Addis, D. R. (2007c). The optimistic brain. *Nat. Neurosci., 10*, 1345–1347.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nat. Rev. Neurosci., 8*, 657–661.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2008). Episodic simulation of future events: Concepts, data, and applications. *The Year in Cognitive Neuroscience, Ann. NY Acad. Sci., 1124*, 39–60.

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annu. Rev. Psychol., 49*, 289–318.

Schacter, D. L., & Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron, 44*, 149–160.

Schacter, D. L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *J. Mem. Lang., 35*, 319–334.

Schnider, A. (2008). *The confabulating mind*. Oxford, UK: Oxford University Press.

Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature, 450*, 102–105.

Slotnick, S. D., & Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nat. Neurosci., 7*, 664–672.

Spreng, R. N., & Levine, B. (2006). The temporal distribution of past and future autobiographical events across the lifespan. *Mem. Cogn., 34*, 1644–1651.

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (in press). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *J. Cogn. Neurosci.*

Stöber, J., & Borkovec, T. D. (2002). Reduced concreteness of worry in generalized anxiety disorder: Findings from a therapy study. *Cogn. Ther. Res., 26*, 89–96.

Suddendorf, T., & Busby, J. (2005). Making decisions with the future in mind: Developmental and comparative identification of mental time travel. *Learn. Motivation, 36*, 110–125.

Suddendorf, T., & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genet. Soc. Gen. Psychol. Monogr., 123*, 133–167.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel and is it unique to humans? *Behav. Brain Sci., 30*, 299–313.

Svoboda, E., Mckinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia*, *44*, 2189–2208.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proc. Natl. Acad. Sci. USA*, *104*, 642–647.

Talland, G. A. (1961). Confabulation in the Wernicke-Korsakoff syndrome. *J. Nerv. Ment. Dis.*, *132*, 361.

Talland, G. A. (1965). *Deranged memory: A psychonomic study of the amnesic syndrome*. New York: Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Clarendon Press.

Tulving, E. (1985). How many memory systems are there? *Am. Psychol.*, *40*, 385–398.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annu. Rev. Psychol.*, *53*, 1–25.

Tulving, E. (2005). Episodic memory and autonoesis. In H. Terrance & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 3–56). New York: Oxford University Press.

Tulving, E., Schacter, D. L., McLachlan, D. R., & Moscovitch, M. (1988). Priming of semantic autobiographical knowledge: A case study of retrograde amnesia. *Brain Cogn.*, *8*, 3–20.

Velakoulis, D., Wood, S. J., Wong, M. T., McGorry, P. D., Yung, A., Phillips, L., et al. (2006). Hippocampal and amygdala volumes according to psychosis stage and diagnosis: A magnetic resonance imaging study of chronic schizophrenia, first-episode psychosis, and ultra-high-risk individuals. *Arch. Gen. Psychiatry*, *63*, 139–149.

Verfaellie, M., Page, K., Orlando, F., & Schacter, D. L. (2005). Impaired implicit memory for gist information in amnesia. *Neuropsychology*, *19*, 760–769.

Vincent, J. L., Snyder, A. Z., Fox, M. D., Shannon, B. J., Andrews, J. R., Raichle, M. E., et al. (2006). Coherent spontaneous activity identifies a hippocampal-parietal memory network. *J. Neurophysiol.*, *96*, 3517–3531.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.*, *9*, 445–453.

Williams, J. M. (1996). Depression and the specificity of autobiographical memory. In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 244–267). Cambridge, UK: Cambridge University Press.

Williams, J. M. (2006). Capture and rumination, functional avoidance, and executive control (CaRFAX): Three processes that underlie overgeneral memory. *Cogn. Emotion*, *20*, 548–568.

Williams, J. M., Ellis, N. C., Tylers, C., Healy, H., Rose, G., & MacLeod, A. K. (1996). The specificity of autobiographical memory and imageability of the future. *Mem. Cogn.*, *24*, 116–125.

VII

LANGUAGE

Chapter 52 HICKOK 767

53 SHAPIRO AND CARAMAZZA 777

54 COHEN AND DEHAENE 789

55 CAPLAN 805

56 HAGOORT, BAGGIO, AND WILLEMS 819

57 KUHL 837

58 RAMUS AND FISHER 855

59 FITCH 873

Introduction

ALFONSO CARAMAZZA

UNTIL FAIRLY RECENTLY almost everything we knew about the cognitive neuroscience of language came to us from investigation of the correlations between patterns of language impairment and their associated loci of brain damage. This approach has contributed importantly to a first-pass characterization of the distribution of language functions in the brain. But its contribution has been even greater to the development of functional theories of language processing. The patterns of dissociations and associations of deficits have informed theories of various language processes from the perception and production of speech, to the organization of the lexicon, to the syntactic and semantic processes involved in sentence comprehension and production. Especially important inroads have been made in characterizing the processing machinery that makes reading and writing possible. Another fertile area of research has been the relation between cognitive and linguistic processes, such as the role of working memory in sentence processing.

Sometimes the observed dissociations and associations of deficits have been rather unexpected, forcing reconsideration of the received view in a given area of language processing. For example, it is now well established that there are patients who make semantic errors in naming pictures orally but can write the name without difficulty, and there are patients with the reverse pattern of dissociation, who make semantic errors in writing the name of an object but who can name orally without problem. These results imply, perhaps unsurprisingly, that the phonological and the orthographic forms of words are represented by distinct neural mechanisms. However, they are not so easily accommodated in theoretical frameworks that hypothesize an abstract lexical representation between the semantic and lexical form levels. Furthermore, since these modality-specific naming deficits have also been found to be restricted to words of one

grammatical class or another, they place even stronger constraints on a biologically defensible theory of the lexicon. Thus, for example, it has been shown that patients may have difficulty in writing verbs but not nouns or speaking verbs but not nouns, while showing no difficulty in the other modality of output for both verbs and nouns (Caramazza & Hillis, 1991). The reverse dissociation, modality-specific selective difficulty with nouns, has also been documented. One implication of these results is that grammatical operations are carried out over modality-specific and not abstract lexical representations. This conclusion is in line with several of the chapters in this section that emphasize the "task-dependent" nature of the representations computed in the course of language processing.

Highly selective dissociations such as those mentioned here place strong constraints on theories of the functional organization of language processes. However, they have been less useful in informing theories of the brain structures that are associated with hypothesized cognitive and linguistic processes. The reason for this discrepancy may be found in the fact that these highly selective deficits are rare and that they are associated with fairly large lesions, making it extremely difficult, if not impossible, to identify the brain regions causally related to the observed dissociations.

Fortunately, the neuropsychological approach is now complemented by various neuroimaging methods that can be used to systematically assess cognitively motivated hypotheses in the intact human brain. This approach is extremely promising but is still in its early stages of development. It has proven difficult to ascribe a causal role to the many brain areas that are activated when performing a complex task. For this reason, it is important to combine the methods of neuropsychology (or transcranial magnetic stimulation), which allow stronger inferences about the causal role of a brain area in the performance of a task, with the increasingly sophisticated use of MRI, EEG, and MEG methods. The chapters in this section fully exemplify the close link between the study of language disorders and neuroimaging research that are being used to converge on a cognitive neuroscience theory of language.

The chapters in this section cover only a limited set of questions about language, reflecting the areas of greatest current interest. The topics include phonological (Hickok) and morphological processing (Shapiro and Caramazza), reading (Cohen and Dehaene), syntactic (Caplan) and semantic processing (Hagoort, Baggio, and Willems), language acquisition (Kuhl), and the genetics (Ramus and Fisher) and evolution of language (Fitch). And, as already noted, in all the chapters that focused on human research, the interplay between the results obtained from the study of language disorders and those obtained from neuroimaging research plays a central role. The result is a comprehensive view of our current understanding of the neurobiology of language.

Consider as an example chapter 54, on reading, which focuses on the word-recognition component of the reading process. Cohen and Dehaene review the various forms of reading impairments that affect some or other aspect of word recognition and conclude that the impairments can be subdivided into two sets: those that principally involve letter recognition and those that involve operations over letters, such as attention mechanisms and serial decoding. They note that the lesions associated with these two broad types of processes concern, respectively, ventral and dorsal visual pathways. They also review imaging results that converge with this interpretation of the neuropsychological evidence. The result is a fairly comprehensive if preliminary view of the neural machinery involved in visual word recognition.

Or consider the case of phonological processing. In chapter 52, Hickok also argues for a two-stream circuit for phonological processing, one dedicated to speech recognition and the other to speech production. Recognition involves primarily a circuit that includes the superior temporal sulcus bilaterally, while production involves a left-dominant frontoparietal/temporal circuit. The evidence for this distinction comes primarily from the patterns of phonological processing deficits in aphasic patients, but Hickok also reviews imaging results that converge with the neuropsychological evidence.

Despite the important developments charted in the chapters included in this section, it is clear that we are still very far from an articulated theory of the biology of language. In some respects we are really only now beginning to develop the methodological and theoretical foundations for such a theory. This fact is illustrated in the beautiful work on the genetics of language described in chapter 58 by Ramus and Fisher, where it is clear that we have only begun to scratch the surface of the many complicated factors that enter in a genetic theory. The same is true for cognitive neuroscience accounts of language, especially for the more complex functions such as syntactic processing and semantic integration. At this stage of the game we have many titillating insights but not yet articulated theories.

REFERENCE

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, *349*, 788–790.

52 The Cortical Organization of Phonological Processing

GREGORY HICKOK

ABSTRACT Phonological processing refers to mechanisms involved in representing, accessing, or manipulating information related to the sound structure of language. The goal of this chapter is to review what is known about the neural basis of phonological processes in three broad domains: speech recognition, speech production, and verbal short-term memory. In particular, we will outline evidence showing that phonological processing is task dependent, that phonological-level aspects of speech recognition are bilaterally organized (but computationally asymmetric), and that posterior phonological information interacts with frontal motor systems by means of a sensory-motor integration network that supports aspects of speech production and verbal working memory. These findings are organized theoretically into a dual-stream model, which is closely related to dual-stream models proposed in the visual domain.

Phonological processing refers to mechanisms involved in representing, accessing, or manipulating information related to the sound structure of language.[1] As such, phonological processes are involved in a range of language abilities. The goal of this chapter is to review what is known about the neural basis of phonological processes in three broad domains: speech recognition, speech production, and verbal short-term memory. We will also examine the relation between these various domains and explore possible parallels and connections between phonological processing networks and cortical systems outside the domain of speech and language.

Phonological processing is task dependent

Given that phonological information is involved in a broad range of linguistic abilities, it is perhaps not surprising that we should find evidence for task dependence in the neural systems recruited to perform this range of tasks. For example, the set of neural circuits involved in, say, verbatim repetition of a heard phonological word form must be at least partially different from the set of neural circuits involved in comprehending the meaning of a heard phonological word form, as the former involves mapping phonological information onto motor-articulatory mechanisms, whereas the later involves mapping phonological information onto lexical-semantic representations. It is an open question whether the phonological *representations* involved in input-related processes, output-related processes, or other processes are shared or distinct—for example, whether there are distinct phonological lexicons (Hickok, 2001; Shelton & Caramazza, 1999)—but it is clear that there are, minimally, distinct and task-dependent *interfaces* that phonological representations enter into (figure 52.1).

A relevant observation regarding task differences in phonological processing comes from a set of studies that were conducted in the late 1970s and early 1980s that showed a double dissociation in two tasks involving phonological processing (Basso, Casati, & Vignolo, 1977; Blumstein, Cooper, Zurif, & Caramazza, 1977; Miceli, Gainotti, Caltagirone, & Masullo, 1980). These studies examined the ability of aphasic patients to perform a syllable discrimination task (e.g., decide whether pairs of syllables such as /da/–/ta/ are the same or different). A prominent theory at the time was that auditory comprehension deficits in aphasia resulted from a deficit in the ability to perceive phonological information in speech (Luria, 1970). Such an account predicted that deficits in syllable discrimination would be strongly associated with auditory comprehension deficits in aphasia. This prediction turned out to be incorrect: a consistent finding was that syllable discrimination and word-level comprehension doubly dissociate, even when the comprehension task involved phonological foils (Baker, Blumstein, & Goodglass, 1981; Miceli et al.). Further, the patient group that tended to perform the worst on syllable discrimination consisted of nonfluent patients with *good* auditory comprehension (Basso et al.). Thus data from aphasia show that it is quite possible to use phonological information to access lexical-semantic information in a comprehension task, yet fail to discriminate syllables, *and* that is it also possible to be able to use phonological information to discriminate syllables, yet fail to comprehend words. See Hickok and Poeppel (2004) for further discussion of these data.

This double dissociation does not imply that there are distinct networks (or lexicons) of phonological *representations*, one involved in syllable discrimination and the other involved

GREGORY HICKOK Center for Cognitive Neuroscience, University of California, Irvine, California

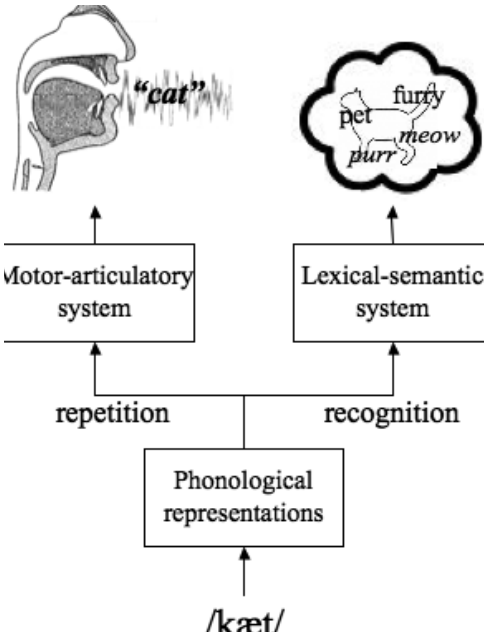


FIGURE 52.1 Distinct mappings from phonological representations to the motor system versus the conceptual-semantic system.

in comprehension. Instead, a more likely explanation is that when these abilities dissociate, the breakdown on the two tasks results from disruption of the task-specific *interface* mechanisms required to perform the respective tasks. For example, auditory comprehension deficits in aphasia appear to arise primarily from postphonemic deficits such as disruption of lexical-semantic access (Baker et al., 1981; Hickok & Poeppel, 2000, 2004, 2007). Similarly, deficits in syllable discrimination (particularly when they occur in nonfluent frontal-lobe-damaged patients with good comprehension) may result from disruption of some component of working memory for phonological information.

The general point is that a deficit on a given phonological task could result from any number of factors that may or may not be task specific. The specific point is that data from syllable discrimination tasks and the like—which are by far the most common means to examine phonological processing—are *not* necessarily measuring the same set of phonological processes involved in auditory comprehension. This is an important point because if we relied on tasks such as syllable discrimination to build our neuroanatomical model, we would conclude that phonological processes are strongly left-hemisphere dominant and are primarily located in the frontal lobe. But if we look instead at phonological processing in the context of auditory comprehension tasks, we arrive at a very different picture, where the system is more bilaterally organized and involves primarily temporal lobe regions (see the next section). Which conclusion is right? The answer is both, and neither. Both are correct in mapping *some* aspect of phonological processing—but clearly different aspects. At the same time, neither is entirely correct because each one alone only paints part of the picture. What is required is the development of a model that can accommodate all sources of data and explain how the involvement of various components of the overall network varies as a function of task.

Spoken word recognition is bilaterally organized

One of the most common, everyday tasks that involve phonological processing is the comprehension of spoken words. According to the dominant theoretical accounts, such as the TRACE (McClelland & Elman, 1986), cohort (Marslen-Wilson, 1987), and neighborhood activation (Luce & Pisoni, 1998) models, spoken word recognition involves several stages of processing, with access to phonological information a critical step in the mapping from sound to meaning. What brain networks support this ability?

Evidence from a variety of sources indicates that spoken word recognition is supported by neural systems in the superior temporal lobe—superior temporal gyrus (STG) and superior temporal sulcus (STS), bilaterally. In neuroimaging studies, a consistent and uncontroversial finding is that, when contrasted with a resting baseline, listening to speech activates the superior temporal lobe bilaterally (Binder et al., 2000, 1994; Mazoyer et al., 1993; Price et al., 1996; Schlosser, Aoyagi, Fulbright, Gore, & McCarthy, 1998; Zatorre, Meyer, Gjedde, & Evans, 1996). It is possible that while activation in spoken word recognition is bilateral, phonological stages of processing are nonetheless restricted to the left hemisphere. This hypothesis predicts that damage to the left posterior superior temporal lobe should produce profound phonological deficits in spoken word recognition. However, this is not the case. Damage to the posterior superior temporal lobe, such as in Wernicke's aphasia (A. Damasio, 1991, 1992) *does* produce deficits in spoken word recognition (Goodglass, 1993; Goodglass, Kaplan, & Barresi, 2001); however, these deficits involve only mild phonological processing impairment and, in fact, appear to result predominantly from a disruption to lexical-semantic-level processes (Bachman & Albert, 1988; Baker et al., 1981; Gainotti, Micelli, Silveri, & Villa, 1982; Miceli et al., 1980). This conclusion is based on experiments in which patients are presented with a spoken word and asked to point to a matching picture within an array that includes phonological, semantic, and unrelated foils; phonological error rates are low overall (5–12%) with semantic errors dominating. This tendency also holds in acute aphasia (Breese & Hillis, 2004; Rogalsky, Pitz, Hillis, & Hickok, 2008) showing that the relative preservation of phonological abilities in unilateral aphasia is not a result of long-term plastic reorganization. Data from split-brain (Zaidel, 1985) and Wada procedures (McGlone, 1984; Hickok et al., 2008) also indicate that the right hemisphere alone is capable of good auditory comprehension at the word level.

In sum, disruption of the left superior temporal lobe does not lead to severe impairments in phonological processing during spoken word recognition. This observation has led to the claim that phonological processes are bilaterally organized in the superior temporal lobe (Hickok & Poeppel, 2000, 2004, 2007). This claim predicts that damage to the STG bilaterally should produce profound impairment in spoken word recognition, which in fact it does in the form of word deafness (Buchman, Garron, Trost-Cardamone, Wichter, & Schwartz, 1986).

The superior temporal sulcus is a critical site for phonological processing

Beyond the earliest stages of speech recognition there is accumulating evidence that portions of the STS are important for representing and/or processing phonological information (Binder et al., 2000; Hickok & Poeppel, 2004, 2007; Indefrey & Levelt, 2004; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Price et al., 1996). The STS is activated by a range of tasks that require access to phonological information, including speech perception and production (Indefrey & Levelt, 2004), and the active short-term maintenance of phonemic information (Buchsbaum, Hickok, & Humphries, 2001; Hickok, Buchsbaum, Humphries, & Muftuler, 2003). Functional imaging studies that attempt to isolate phonological processes in perception by contrasting speech stimuli with complex nonspeech signals have found activation along the STS (Liebenthal et al., 2005; Narain et al., 2003; Obleser, Zimmermann, Van Meter, & Rauschecker, 2006; Scott, Blank, Rosen, & Wise, 2000; Spitsyna, Warren, Scott, Turkheimer, & Wise, 2006; Vouloumanos, Kiehl, Werker, & Liddle, 2001), as have studies that manipulate psycholinguistic variables that tap phonological networks (Okada & Hickok, 2006), such as phonological neighborhood density (the number of words that sound similar to a target word). Although many authors consider this system to be strongly left dominant, both lesion evidence and imaging evidence suggest a bilateral organization (Hickok & Poeppel, 2007).

One currently unresolved question is the relative contribution of anterior versus posterior STS regions in phonological processing. Lesion evidence indicates that damage to posterior temporal lobe areas is most predictive of auditory comprehension deficits (Bates et al., 2003); however, as noted earlier, comprehension deficits in aphasia result predominantly from postphonemic processing levels. A majority of functional imaging studies targeting phonological processing in perception have highlighted regions in the posterior half of the STS (Hickok & Poeppel, 2007). Other studies, however, have reported *anterior* STS activation in perceptual speech tasks (Mazoyer et al., 1993; Narain et al., 2003; Scott et al., 2000; Spitsyna et al., 2006). These studies involved sentence-level stimuli, raising the possibility that anterior STS regions may be responding to some other aspect of the stimuli such as its syntactic or prosodic organization (Friederici, Meyer, & von Cramon, 2000; Humphries, Binder, Medler, & Liebenthal, 2006; Humphries, Love, Swinney, & Hickok, 2005; Humphries, Willard, Buchsbaum, & Hickok, 2001; Vandenberghe, Nobre, & Price, 2002). The weight of the available evidence, therefore, suggests that the critical portion of the STS that is involved in phonological-level processes is bounded anteriorly by the anterolateralmost aspect of Heschl's gyrus and posteriorly by the posteriormost extent of the Sylvian fissure (Hickok & Poeppel, 2007).

Phonological processing systems in speech recognition are bilateral but asymmetric

The claim that phonological processing is bilaterally organized for speech recognition tasks does not imply that the systems in both hemispheres are computationally identical. To the contrary, there is abundant evidence for differences in the way acoustic/speech information is processed in the two hemispheres (Abrams, Nicol, Zecker, & Kraus, 2008; Boemio, Fromm, Braun, & Poeppel, 2005; Giraud et al., 2007; Hickok & Poeppel, 2007; Zatorre, Belin, & Penhune, 2002). What is less clear is the computational nature of these differences. One view is that the difference turns on biases toward temporal (left-hemisphere) versus spectral (right-hemisphere) resolution (Zatorre et al., 2002). Another view holds that the two hemispheres differ in terms of their sampling rate, with the left hemisphere operating at a higher rate (25–50 Hz) and the right hemisphere at a lower rate (3–5 Hz) (Poeppel, 2003).[2] Yet another proposal, more specific to phonological processing, is that the left hemisphere processes phonemic information in a categorical fashion, whereas the right hemisphere may treat such information in a more continuous fashion (Liebenthal et al., 2005). We will not resolve these questions here. For our purposes, it is important to note that computational differences exist between the two hemispheres in the way that speech signals are processed during speech recognition, but that both are involved in the process, and both are largely capable of processing phonological information sufficiently well to access lexical-semantic information (Hickok & Poeppel, 2004). This analysis indicates that spoken word recognition involves parallel pathways (multiple routes) in the mapping from sound to meaning (Hickok & Poeppel, 2007). Although this conclusion differs from standard models of speech recognition (Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland & Elman, 1986), it agrees nicely with the fact that speech contains redundant cues to phonemic information, as well as with behavioral evidence suggesting that the speech system can take advantage of these different cues

(Remez, Rubin, Pisoni, & Carrell, 1981; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

Posterior language cortex in the left hemisphere is involved in phonological aspects of speech production

There is unequivocal evidence that posterior sensory-related cortex in the left, but not right, hemisphere plays an important role in speech production. For example, damage to the left posterior temporal lobe often results not only in comprehension deficits, but also in speech *production* deficits (A. Damasio, 1992; H. Damasio, 1991; Geschwind, 1971; Goodglass, 1993; Goodglass et al., 2001). Disruption to phonological systems appears to account for some of these production deficits. Damage to the left dorsal STG and/or the supramarginal gyrus/temporal-parietal junction is associated with conduction aphasia, a syndrome that is characterized by good comprehension, but with frequent phonemic errors in speech production, naming difficulties that often involve tip-of-the-tongue states (implicating a breakdown in phonological encoding), and difficulty with verbatim repetition (H. Damasio & Damasio, 1980; Goodglass, 1992).[3] Conduction aphasia has classically been considered to be a disconnection syndrome involving damage to the arcuate fasciculus (Geschwind, 1965). However, there is now good evidence that this syndrome results from cortical dysfunction (Anderson et al., 1999; Hickok et al., 2000). The production deficit is load sensitive: errors are more likely on longer, lower-frequency words and verbatim repetition of strings of speech with little semantic constraint (Goodglass, 1992, 1993). Functionally, conduction aphasia has been characterized as a deficit in the ability to encode phonological information for production (Wilshire & McCarthy, 1996). Thus conduction aphasia provides evidence for the involvement of left posterior auditory-related brain regions in the phonological aspect of speech production. See also Wise et al. (2001).

Functional imaging evidence also implicates left superior posterior temporal regions in speech production generally (Hickok et al., 2000; Price et al., 1996) and phonological stages of the process in particular (Indefrey & Levelt, 2000, 2004). With respect to the latter, the posterior portion of the left planum temporale region, which is within the distribution of lesions associated with conduction aphasia, activates during picture naming and exhibits length effects (Okada, Smith, Humphries, & Hickok, 2003) and frequency effects (Graves, Grabowski, Mahta, & Gordon, 2007), and has a time course of activation, measured electromagnetically, that is consistent with the phonological encoding stage of naming (Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998).

Taken together, the lesion evidence and physiological evidence reviewed in this section make a compelling argument for the involvement of left posterior superior temporal regions in phonological aspects of speech production.

The left posterior planum temporale is part of an auditory-motor integration circuit

If left posterior superior temporal regions are involved in phonological aspects of speech *production*, there must be a mechanism for interfacing posterior and anterior brain regions. The need for such a mechanism has long been acknowledged, and in classical models was instantiated as a simple white matter pathway, the arcuate fasciculus (Geschwind, 1971). More recent proposals have argued, instead, for a *cortical* system that serves to integrate sensory and motor aspects of speech (Hickok et al., 2000, 2003; Hickok & Poeppel, 2000, 2004, 2007; Warren, Wise, & Warren, 2005), which is consistent with much research on sensorimotor integration systems studied in the context of the monkey visual system (Andersen, 1997; Colby & Goldberg, 1999; Milner & Goodale, 1995).

A series of studies over the last several years have identified a cortical network for speech and related abilities (e.g., music), which has many of the properties exhibited by sensorimotor networks studied in other domains. These properties include sensorimotor responses, connectivity with frontal motor systems, motor-effector specificity, and multisensory responses (Andersen, 1997; Colby & Goldberg, 1999). The speech-related network with these response properties includes an area (termed Spt) in the left posterior planum temporale region, that has been argued to support sensorimotor integration for speech (Hickok et al., 2003). We will review the evidence for this claim in the following paragraphs.

Spt Exhibits Sensorimotor Response Properties A number of studies have demonstrated the existence of an area in the left posterior planum temporale that responds during both the perception and production of speech, even when speech is produced covertly (subvocally) so that there is no overt auditory feedback (Buchsbaum et al., 2001; Buchsbaum, Olsen, Koch, & Berman, 2005; Buchsbaum, Olsen, Koch, Kohn, et al., 2005; Hickok et al., 2003). Spt is not speech specific, however. It responds equally well to the perception and (covert) production by humming of melodic stimuli (Hickok et al.; Pa & Hickok, 2008).

Spt Is Functionally Connected to Motor Speech Areas Spt activity is tightly correlated with activity in frontal speech-production-related areas, such as the pars opercularis (BA 44) (Buchsbaum et al., 2001), suggesting that the two regions are functionally connected. Furthermore, cortex in the posterior portion of the planum temporale (area Tpt) has a cytoarchitectonic structure that is similar to BA44. Galaburda writes that area Tpt "exhibits a degree of specialization like that of Area 44 in Broca's region. It contains prominent pyramids in layer IIIc and a broad

lamina IV. . . . the intimate relationship and similar evolutionary status of Areas 44 and Tpt allows for a certain functional overlap" (Galaburda, 1982, p. 442).

Spt Activity Is Modulated by Motor Effector Manipulations In monkey parietal cortex, sensorimotor integration areas are organized around motor effector systems (e.g., ocular versus manual actions in LIP and AIP; Andersen, 1997; Colby & Goldberg, 1999). Recent evidence suggests that Spt may be organized around the vocal tract effector system: Spt was less active when skilled pianists listened to and then imagined playing novel melodies than when they listened to and covertly hummed the same melodies (Pa & Hickok, 2008).

Spt Is Sensitive to Speech-Related Visual Stimuli Many neurons in sensorimotor integration areas of the monkey parietal cortex are sensitive to inputs from more than one sensory modality (Andersen, 1997). The planum temporale, while often thought to be an auditory area, also activates in response to sensory input from other modalities. For example, silent lipreading has been shown to activate auditory cortex in the vicinity of the planum temporale (Calvert et al., 1997; Calvert & Campbell, 2003). Although these studies typically report the location as "auditory cortex" including primary regions, group-based localizations in this region can be unreliable. Indeed, a recent fMRI study using individual subject analyses has found that activation to visual speech and activation using the standard Spt-defining auditory-motor task (listen then covertly produce) are found in the same regions of the left posterior planum temporale (Okada & Hickok, 2009). Thus Spt appears to be sensitive also to visual input that is relevant to vocal tract actions.

In summary, Spt exhibits all the features of sensorimotor integration areas as identified in the parietal cortex of the monkey. This finding suggests that Spt is a sensorimotor integration area for vocal tract actions (Pa & Hickok, 2008), placing it in the context of a network of sensorimotor integration areas in the posterior parietal and temporal/parietal cortex, which receive multisensory input and are organized around motor-effector systems (Andersen, 1997). Although area Spt is not language specific, it counts sensorimotor integration for phonological information as a prominent function.

Verbal short-term memory relies on auditory-motor integration networks

Verbal short-term memory is often held to comprise at least two components: a storage component of some form and a mechanism for active maintenance of this information. In Baddeley's model, for example, the storage mechanism is the "phonological store," a dedicated buffer, and active maintenance is achieved by the "articulatory rehearsal" mechanism (Baddeley, 1992). The concept of a sensorimotor integration network, as outlined previously, provides an independently motivated neural circuit that may be the basis for verbal short-term memory (Buchsbaum, Olsen, Koch, & Berman, 2005; Hickok et al., 2003; Hickok & Poeppel, 2000; see also Aboitiz & García V., 1997; Jacquemot & Scott, 2006). Specifically, on the assumption that the proposed sensorimotor integration circuit is bidirectional (Hickok & Poeppel, 2000, 2004, 2007), one can equate the storage component of verbal short-term memory with sensory representations in the superior temporal lobe (the same STS regions that are involved in sensory/recognition processes), and one can equate the active maintenance component with frontal articulatory systems: the sensorimotor integration network (Spt) allows articulatory mechanisms to maintain verbal information in an active state (Hickok et al., 2003). In this sense, the basic architecture is similar to Baddeley's, except that there is a proposed computational mechanism (sensorimotor transformations in Spt) mediating the relation between the storage and active maintenance components. This view differs from Baddeley's, however, in that it assumes that the storage component is not a dedicated buffer but an active state of networks that are involved in perceptual recognition (Fuster, 1995; Ruchkin, Grafman, Cameron, & Berndt, 2003). Because our evidence suggests that the proposed sensorimotor integration network is not specific to phonological information (Hickok et al.), we also suggest that the verbal short-term memory circuit is not specific to phonological information, a position that is in line with recent behavioral work (Jones, Hughes, & Macken, 2007; Jones & Macken, 1996; Jones, Macken, & Nicholls, 2004). For a thorough discussion of these issues, see Buchsbaum and D'Esposito (2008).

A theoretical framework: The dual-stream model

The processing of phonological information in speech recognition, speech production, and short-term memory involves partially overlapping, but also partially distinct, neural circuits. Speech recognition relies primarily on neural circuits in the superior temporal lobes bilaterally, whereas speech production and verbal short-term memory rely on a frontoparietal/temporal circuit that is left-hemisphere dominant. As noted earlier, this divergence of processing streams is consistent with the fact that phonological information plays a role in (1) accessing lexical-semantic representations on the one hand and (2) driving motor-speech articulation on the other. As lexical-semantic and motor-speech systems involve very different types of representations and processing mechanisms, it stands to reason that divergent pathways underlie the interface with phonological networks.

The dual-interface requirements with respect to phonological processing are captured neuroanatomically by the dual-stream model[4] (figure 52.2) (Hickok & Poeppel, 2000, 2004, 2007). The model is rooted in dual-stream proposals in vision (Milner & Goodale, 1995) that distinguish between a ventral stream involved in visual object recognition ("what" stream) and a dorsal stream involved in visual-motor integration (sometimes called a "how" stream). Accordingly, the dual-stream model proposes that a ventral stream, which involves structures in the superior and middle portions of the temporal lobe, is involved in processing speech signals for comprehension (speech recognition), whereas a dorsal stream, which involves structures in the posterior dorsalmost aspect of the temporal lobe and parietal operculum, as well as the posterior frontal lobe, is involved in translating speech signals into articulatory representations in the frontal lobe. The suggestion that the dorsal stream has an auditory-motor integration function differs from earlier arguments for a dorsal auditory "where" system (Rauschecker, 1998) but has gained support in recent years (Scott & Johnsrude, 2003; Warren et al., 2005; Wise et al., 2001).

The dual-stream model can explain the double dissociations between syllable discrimination tasks and auditory comprehension tasks noted earlier on the assumption

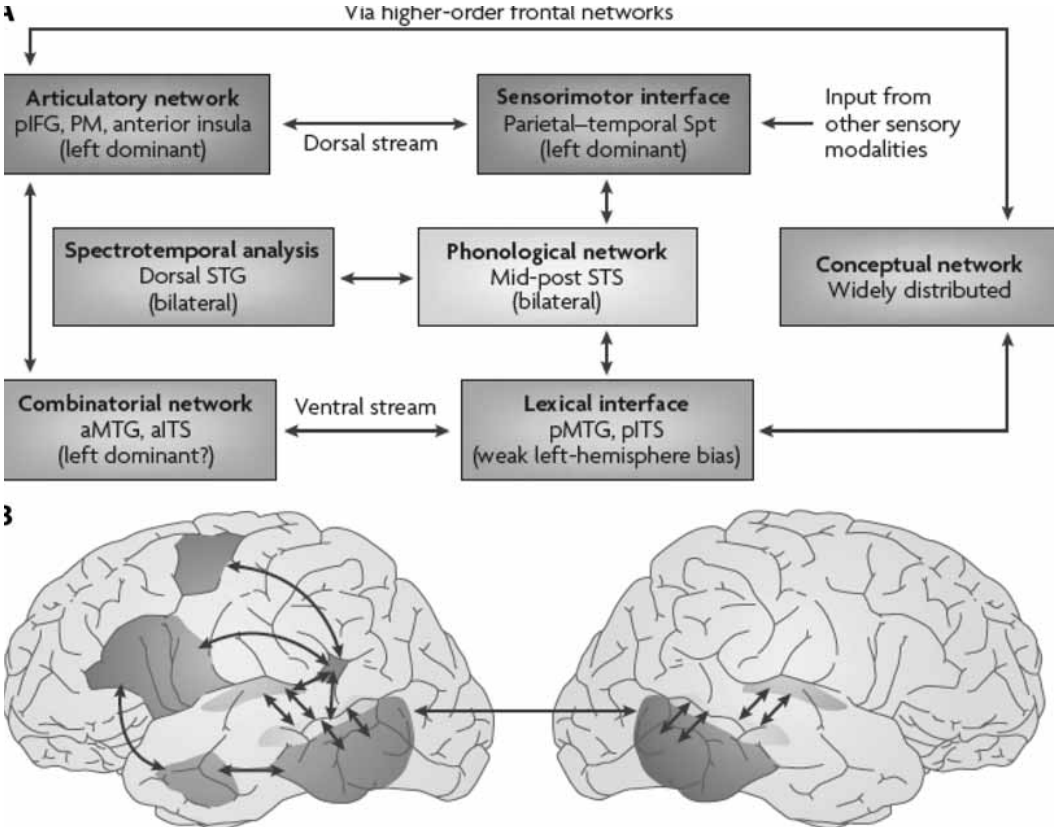


FIGURE 52.2 The dual-stream model of speech processing. (*A*) Schematic diagram of the dual-stream model. The earliest stage of cortical speech processing involves some form of spectrotemporal analysis, which is carried out in auditory cortices bilaterally in the supratemporal plane. These spectrotemporal computations appear to differ between the two hemispheres. Phonological-level processing and representation involves the middle to posterior portions of superior temporal sulcus (STS) bilaterally, although there may be a weak left-hemisphere bias at this level of processing. Subsequently, the system diverges into two broad streams, a dorsal pathway (blue) that maps sensory or phonological representations onto articulatory motor representations, and a ventral pathway (red) that maps sensory or phonological representations onto lexical-conceptual representations. (*B*) Approximate anatomical locations of the dual-stream model components, specified as precisely as available evidence allows. Regions shaded green depict areas on the dorsal surface of the STG that are hypothesized to be involved in spectrotemporal analysis. Regions shaded yellow in the posterior half of the STS are implicated in phonological-level processes. Regions shaded red represent the ventral stream, which is bilaterally organized with a weak left-hemisphere bias. The more posterior regions of the ventral stream, the posterior middle and inferior portions of the temporal lobes, correspond to the *lexical interface*, which links phonological and semantic information, whereas the more anterior locations correspond to the hypothesized *combinatorial network*. Regions shaded blue represent the dorsal stream, which is strongly left-dominant. The posterior region of the dorsal stream corresponds to an area in the Sylvian fissure at the parietal-temporal boundary (area Spt), which is hypothesized to be a *sensorimotor interface*, whereas the more anterior locations in the frontal lobe, likely involving Broca's region and a more dorsal premotor site, correspond to portions of the articulatory network. (Figure reproduced from Hickok & Poeppel, 2007.) (See color plate 64.)

that syllable discrimination relies to a greater extent on dorsal stream circuitry (Burton, Small, & Blumstein, 2000) (explaining the association with frontal lesions), whereas speech recognition tasks rely to a greater extent on ventral stream circuitry. The involvement of dorsal stream circuitry in syllable discrimination tasks makes sense given that discrimination of serially presented speech information requires some degree of verbal short-term memory. In addition, in contrast to the typical view that speech processing is mainly left-hemisphere dependent, the model suggests that the ventral stream is bilaterally organized (although with important computational differences between the two hemispheres); thus the ventral stream itself comprises parallel processing streams. This approach would explain the failure to find substantial speech recognition deficits following unilateral temporal lobe damage. The dorsal stream, however, is strongly left-dominant, explaining why production deficits are prominent sequelae of dorsal temporal and frontal lesions, as well as explaining why left-hemisphere injury can substantially impair performance on syllable discrimination tasks (Hickok & Poeppel, 2000, 2004, 2007).

On mirror neurons and motor theories of perception

Evidence we have reviewed suggests a tight connection between systems involved in speech perception and speech production, and the dual-stream model captures this association in the form of the dorsal processing stream that mediates this relation. The idea that perception and production systems in speech are functionally interrelated is not new, as it was an integral component of Wernicke's language model of 1874 (Wernicke, 1874/1969). The motor theory of speech perception also highlighted important links between perception and production, but with quite a different spin. Whereas Wernicke emphasized the role of perceptual systems in guiding speech production, the motor theory proposed the reverse, that motor speech systems were the foundation for speech perception (Liberman & Mattingly, 1985). Although the motor theory had lost favor among most speech/language scientists, the discovery of "mirror neurons" (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) has triggered a resurgence of interest in motor theories of perception generally (Iacoboni et al., 2005; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), and the motor theory of speech perception in particular (Rizzolatti & Arbib, 1998). Mirror neurons are cells found in monkey frontal cortex that respond both during the execution of motor acts and during the perception of others performing similar motor acts. It has been suggested that mirror neurons are the basis for action understanding, including the perception/understanding of speech (Rizzolatti & Arbib, 1998).

Despite the current popularity of motor theories of speech perception, there is strong evidence that the theory is incorrect. Motor theories of speech perception make a clear prediction: disruption of the motor systems involved in speech production should produce a substantial disruption of speech recognition. This prediction is falsified by the common occurrence of patients with large left frontal lesions who have profound impairments in the ability to produce speech, yet have well-preserved ability to comprehend speech at the lexical level (i.e., severe Broca's aphasics) (Goodglass, 1993; Goodglass et al., 2001). This finding demonstrates clearly that speech recognition can be achieved without the motor-speech system's involvement. However, damage to sensory-related speech areas regularly produces deficits in speech production, such as the paraphasic errors found in the fluent speech of Wernicke and conduction aphasics (Goodglass, 1993; Goodglass et al., 2001). Thus the evidence confirms Wernicke's conceptualization of the relation between sensory and motor speech systems, namely, that sensory systems are necessary for speech production, but motor-speech systems are not necessary for speech recognition (Hickok & Poeppel, 2000, 2004, 2007). Put differently, the relation between sensory and motor speech systems is better characterized by a *perceptual theory of speech production* than a *motor theory of speech perception*.

A strong version of a motor/mirror-neuron theory of speech perception is clearly untenable. At the same time, it is quite clear that motor knowledge can influence perception (Galantucci, Fowler, & Turvey, 2006), as the McGurk effect clearly demonstrates (McGurk & MacDonald, 1976). These effects do not imply, however, that speech perception requires the involvement of motor systems, only that motor knowledge can influence or constrain the acoustic analysis of speech, for example, by means of top-down, or predictive, coding mechanisms (van Wassenhove, Grant, & Poeppel, 2005). The proposed sensorimotor integration network provides a neural basis for this influence of motor knowledge on speech perception.

Summary

Phonological processing is a heterogeneous, task-dependent construct, and the neural systems that support phonological processing are similarly heterogeneous and task dependent. There is a fundamental distinction between the processes and neural circuits involved in tasks that involve motor-related systems compared with tasks that primarily involve lexical-semantic systems leading to task-related double dissociations within the context of "phonological processing." There is also a substantial amount of interaction between sensory- and motor-related aspects of phonological processing, as well as evidence for shared resources, such as phonological systems in the STS. The neuroanatomical framework

provided by the dual-stream model captures the distinctions between phonological tasks, as well as provides a basis for sensorimotor interactions.

ACKNOWLEDGMENTS This work was supported by NIH grant DC03681.

NOTES

1. Similar processes appear to operate in visual-manual languages (signed languages) suggesting that the association with sound per se may not be a defining feature of phonology (Emmorey, 2002).
2. These two proposals may not be incompatible as there is a relation between sampling rate and spectral versus temporal resolution (Zatorre, Belin, & Penhune, 2002).
3. Although conduction aphasia is often characterized as a disorder of repetition, it is clear that the deficit extends well beyond this one task (Hickok et al., 2000). In fact, Wernicke first identified conduction aphasia as a disorder of speech production in the face of preserved comprehension (Wernicke, 1874/1969). It was only later that Lichtheim introduced repetition as a convenient diagnostic tool for assessing the integrity of the link between sensory and motor speech systems (Lichtheim, 1885).
4. Note that this model is not intended as a psycholinguistic model. Nor are the boxes/brain areas intended to correspond to boxes in any existing psycholinguistic model of speech production or recognition. Instead it is a neuroanatomical outline of the brain regions involved in some linguistic operations, such as speech recognition/comprehension, speech production, and phonological working memory.

REFERENCES

ABOITIZ, F., & GARCÍA, V. R. (1997). The evolutionary origin of language areas in the human brain: A neuroanatomical perspective. *Brain Res. Brain Res. Rev.*, *25*, 381–396.

ABRAMS, D. A., NICOL, T., ZECKER, S., & KRAUS, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.*, *28*(15), 3958–3965.

ANDERSEN, R. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *352*, 1421–1428.

ANDERSON, J. M., GILMORE, R., ROPER, S., CROSSON, B., BAUER, R. M., NADEAU, S., BEVERSDORF, D. Q., CIBULA, J., ROGISH, M., III, KORTENCAMP, S., HUGHES, J. D., GONZALEZ ROTHI, L. J., & HEILMAN, K. M. (1999). Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. *Brain Lang.*, *70*, 1–12.

BACHMAN, D. L., & ALBERT, M. L. (1988). Auditory comprehension in aphasia. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 1, pp. 281–306). New York: Elsevier.

BADDELEY, A. D. (1992). Working memory. *Science, 255*, 556–559.

BAKER, E., BLUMSTEIN, S. E., & GOODGLASS, H. (1981). Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia*, *19*, 1–15.

BASSO, A., CASATI, G., & VIGNOLO, L. A. (1977). Phonemic identification defects in aphasia. *Cortex*, *13*, 84–95.

BATES, E., WILSON, S. M., SAYGIN, A. P., DICK, F., SERENO, M. I., KNIGHT, R. T., & DRONKERS, N. F. (2003). Voxel-based lesion-symptom mapping. *Nat. Neurosci.*, *6*(5), 448–450.

BINDER, J. R., FROST, J. A., HAMMEKE, T. A., BELLGOWAN, P. S., SPRINGER, J. A., KAUFMAN, J. N., & POSSING, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex*, *10*, 512–528.

BINDER, J. R., RAO, S. M., HAMMEKE, T. A., YETKIN, F. Z., JESMANOWICZ, A., BANDETTINI, P. A., WONG, E. C., ESTKOWSKI, L. D., GOLDSTEIN, M. D., HAUGHTON, V. M., & HYDE, J. S. (1994). Functional magnetic resonance imaging of human auditory cortex. *Ann. Neurol.*, *35*, 662–672.

BLUMSTEIN, S. E., COOPER, W. E., ZURIF, E. B., & CARAMAZZA, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia*, *15*, 371–383.

BOEMIO, A., FROMM, S., BRAUN, A., & POEPPEL, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.*, *8*(3), 389–395.

BREESE, E. L., & HILLIS, A. E. (2004). Auditory comprehension: Is multiple choice really good enough? *Brain Lang.*, *89*(1), 3–8.

BUCHMAN, A. S., GARRON, D. C., TROST-CARDAMONE, J. E., WICHTER, M. D., & SCHWARTZ, M. (1986). Word deafness: One hundred years later. *J. Neurol. Neurosurg. Psychiatry*, *49*, 489–499.

BUCHSBAUM, B. R., & D'ESPOSITO, M. (2008). The search for the phonological store: From loop to convolution. *J. Cogn. Neurosci.*, *20*(5), 762–778.

BUCHSBAUM, B., HICKOK, G., & HUMPHRIES, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.*, *25*, 663–678.

BUCHSBAUM, B. R., OLSEN, R. K., KOCH, P., & BERMAN, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, *48*(4), 687–697.

BUCHSBAUM, B. R., OLSEN, R. K., KOCH, P. F., KOHN, P., KIPPENHAN, J. S., & BERMAN, K. F. (2005). Reading, hearing, and the planum temporale. *NeuroImage*, *24*(2), 444–454.

BURTON, M. W., SMALL, S., & BLUMSTEIN, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *J. Cogn. Neurosci.*, *12*, 679–690.

CALVERT, G. A., BULLMORE, E. T., BRAMMER, M. J., CAMPBELL, R., WILLIAMS, S. C. R., MCGUIRE, P. K., WOODRUFF, P. W. R., IVERSEN, S. D., & DAVID, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*, 593–596.

CALVERT, G. A., & CAMPBELL, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *J. Cogn. Neurosci.*, *15*, 57–70.

COLBY, C. L., & GOLDBERG, M. E. (1999). Space and attention in parietal cortex. *Annu. Rev. Neurosci.*, *22*, 319–349.

DAMASIO, A. R. (1991). Signs of aphasia. In M. T. Sarno (Ed.), *Acquired aphasia* (2nd ed., pp. 27–43). San Diego: Academic Press.

DAMASIO, A. R. (1992). Aphasia. *N. Engl. J. Med.*, *326*, 531–539.

DAMASIO, H. (1991). Neuroanatomical correlates of the aphasias. In M. Sarno (Ed.), *Acquired aphasia* (2nd ed., pp. 45–71). San Diego: Academic Press.

DAMASIO, H., & DAMASIO, A. R. (1980). The anatomical basis of conduction aphasia. *Brain*, *103*, 337–350.

DI PELLEGRINO, G., FADIGA, L., FOGASSI, L., GALLESE, V., & RIZZOLATTI, G. (1992). Understanding motor events: A neurophysiological study. *Exp. Brain Res.*, *91*(1), 176–180.

EMMOREY, K. (2002). *Language, cognition, and the brain: Insights from sign language research.* Mahwah, NJ: Lawrence Erlbaum.

FRIEDERICI, A. D., MEYER, M., & VON CRAMON, D. Y. (2000). Auditory languge comprehension: An event-related fMRI study on

the processing of syntactic and lexical information. *Brain Lang.*, *74*, 289–300.

Fuster, J. M. (1995). *Memory in the cerebral cortex.* Cambridge, MA: MIT Press.

Gainotti, G., Micelli, G., Silveri, M. C., & Villa, G. (1982). Some anatomo-clinical aspects of phonemic and semantic comprehension disorders in aphasia. *Acta Neurol. Scand.*, *66*, 652–665.

Galaburda, A. M. (1982). Histology, architectonics, and asymmetry of language areas. In M. A. Arbib, D. Caplan, & J. C. Marshall (Eds.), *Neural models of language processes* (pp. 435–445). San Diego: Academic Press.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.*, *13*(3), 361–377.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(Pt. 2), 593–609.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain*, *88*, 237–294, 585–644.

Geschwind, N. (1971). Aphasia. *N. Engl. J. Med.*, *284*, 654–656.

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, *56*(6), 1127–1134.

Goodglass, H. (1992). Diagnosis of conduction aphasia. In S. E. Kohn (Ed.), *Conduction aphasia* (pp. 39–49). Hillsdale, NJ: Lawrence Erlbaum.

Goodglass, H. (1993). *Understanding aphasia.* San Diego: Academic Press.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The assessment of aphasia and related disorders* (3rd ed.). Philadelphia: Lippincott Williams & Wilkins.

Graves, W. W., Grabowski, T. J., Mahta, S., & Gordon, J. K. (2007). A neural signature of phonological access: Distinguishing the effects of word frequency from familiarity and length in overt picture naming. *J. Cogn. Neurosci.*, *19*, 617–631.

Hickok, G. (2001). Functional anatomy of speech perception and speech production: Psycholinguistic implications. *J. Psycholinguist. Res.*, *30*, 225–234.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *J. Cogn. Neurosci.*, *15*, 673–682.

Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A. K., Naeve-Velguth, S., Strupp, J. P., Strick, P. L., & Ugurbil, K. (2000). A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: Implications for the explanation of conduction aphasia. *Neurosci. Lett.*, *287*, 156–160.

Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., Barde, L., & Grant, A. (2008). Bilateral capacity for speech sound processing in auditory comprehension: Evidence from Wada procedures. *Brain Lang.*, *107*(3), 179–184.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.*, *4*, 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.*, *8*(5), 393–402.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J. Cogn. Neurosci.*, *18*(4), 665–679.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum. Brain Mapp.*, *26*, 128–138.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *NeuroReport*, *12*, 1749–1752.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.*, *3*(3), e79.

Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 845–865). Cambridge, MA: MIT Press.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.*, *10*, 480–486.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *Q. J. Exp. Psychol.*, *60*(4), 505–511.

Jones, D. M., & Macken, W. J. (1996). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.*, *19*, 369–381.

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *J. Exp. Psychol. Learn. Mem. Cogn.*, *30*(3), 656–674.

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *J. Cogn. Neurosci.*, *10*, 553–567.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Lichtheim, L. (1885). On aphasia. *Brain*, *7*, 433–484.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cereb. Cortex*, *15*(10), 1621–1631.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear.*, *19*, 1–36.

Luria, A. R. (1970). *Traumatic aphasia.* The Hague: Mouton.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. *J. Cogn. Neurosci.*, *5*, 467–479.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.*, *18*, 1–86.

McGlone, J. (1984). Speech comprehension after unilateral injection of sodium amytal. *Brain Lang.*, *22*, 150–157.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Miceli, G., Gainotti, G., Caltagirone, C., & Masullo, C. (1980). Some aspects of phonological impairment in aphasia. *Brain Lang.*, *11*, 159–169.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action.* Oxford, UK: Oxford University Press.

Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a left-lateralized

response specific to intelligible speech using fMRI. *Cereb. Cortex*, *13*(12), 1362–1368.

Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2006). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb. Cortex*, *17*, 2251–2257.

Okada, K., & Hickok, G. (2006). Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. *NeuroReport*, *17*, 1293–1296.

Okada, K., & Hickok, G. (2008). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neurosci. Lett.* [Epub ahead of print. doi: 10.1016/j.neulet.2009.01.060.]

Okada, K., Smith, K. R., Humphries, C., & Hickok, G. (2003). Word length modulates neural activity in auditory cortex during covert object naming. *NeuroReport*, *14*, 2323–2326.

Pa, J., & Hickok, G. (2008). A parietal-temporal sensorymotor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. *Neuropsychologia*, *46*, 362–368.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Commun.*, *41*, 245–255.

Price, C. J., Wise, R. J. S., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., Frackowiak, R. S. J., & Friston, K. J. (1996). Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain*, *119*, 919–931.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Curr. Opin. Neurobiol.*, *8*(4), 516–521.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*, 947–950.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends Neurosci.*, *21*, 188–194.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.*, *3*(2), 131–141.

Rogalsky, C., Pitz, E., Hillis, A. E., & Hickok, G. (2008). Auditory word comprehension impairment in acute stroke: Relative contribution of phonemic versus semantic factors. *Brain Lang.*, *107*(2), 167–169.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behav. Brain Sci.*, *26*, 709–777.

Schlosser, M. J., Aoyagi, N., Fulbright, R. K., Gore, J. C., & McCarthy, G. (1998). Functional MRI studies of auditory comprehension. *Hum. Brain Mapp.*, *6*, 1–13.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–2406.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.*, *26*(2), 100–107.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Shelton, J. R., & Caramazza, A. (1999). Deficits in lexical and semantic processing: Implications for models of normal language. *Psychon. Bull. Rev.*, *6*, 5–27.

Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2006). Converging language streams in the human temporal lobe. *J. Neurosci.*, *26*(28), 7328–7336.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. USA*, *102*(4), 1181–1186.

Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *J. Cogn. Neurosci.*, *14*(4), 550–560.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *J. Cogn. Neurosci.*, *13*(7), 994–1005.

Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends Neurosci.*, *28*(12), 636–643.

Wernicke, C. (1874/1969). The symptom complex of aphasia: A psychological study on an anatomical basis. In R. S. Cohen & M. W. Wartofsky (Eds.), *Boston studies in the philosophy of science* (pp. 34–97). Dordrecht: D. Reidel.

Wilshire, C. E., & McCarthy, R. A. (1996). Experimental investigations of an impairement in phonological encoding. *Cogn. Neuropsychol.*, *13*, 1059–1098.

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within "Wernicke's area." *Brain*, *124*, 83–95.

Zaidel, E. (1985). Language in the right hemisphere. In D. F. Benson & E. Zaidel (Eds.), *The dual brain: Hemispheric specialization in humans* (pp. 205–231). New York: Guilford Press.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends Cogn. Sci.*, *6*, 37–46.

Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: Review, replication, and reanalysis. *Cereb. Cortex*, *6*, 21–30.

53 Morphological Processes in Language Production

KEVIN A. SHAPIRO AND ALFONSO CARAMAZZA

ABSTRACT Morphology refers to the set of linguistic processes that govern the composition of words from stored units called morphemes, which encode information about meaning and grammatical properties. Neuropsychological studies suggest that morphological operations can be spared or impaired in the setting of acquired brain damage. Moreover, specific patterns of breakdown in morphology have revealed major principles underlying the neural architecture of language. Here we make the case that the language production system has at least three components with discrete neural substrates: one component that represents lexical concepts and is organized according to meaning; a second component that processes morphological information linked to grammatical function; and a third component that converts lexical and morphological representations into specific output forms.

The basic unit of meaning in language is the morpheme, a type of cognitive representation that corresponds either to a lexical concept (a root, like *think*), to an abstract modifier that can be used to generate new lexical concepts with distinct meanings (a derivational morpheme, like *re-* or *-able*), or to a property relevant to the grammatical rules of a language (a functional morpheme, like the preposition *of* or the past tense marker *-ed*). Morphology—the system of rules that governs the construction of words from individual morphemes—is the engine that drives much of language's combinatorial productivity, bridging the gap between the conceptual, grammatical, and phonological levels of representation. Generative morphological rules are also extremely versatile, allowing speakers to express practically unlimited nuances of meaning (*unthinkable, redirected, antidisestablishmentarians,* etc.) using a fixed set of stored representational elements.

Languages differ widely in the way that morphological structure is realized in the phonological message. In English and many other languages, morphemes may be either phonologically unbound, in the sense that they can be produced separately from other morphemes (for example, prepositions), or bound, meaning that they cannot be produced in isolation (like the markers of plural number and past tense). Bound functional morphemes are called inflections. In Mandarin Chinese and other so-called isolating languages, functional morphemes are, as a rule, unbound; for example, the perfective aspect marker *le* in the sentence *wǒ mǎi le sānběn shū* ("I bought three books") indicates that the action expressed by the verb *mǎi* (buying) has been completed.

This difference in phonological expression should not be taken to imply that English is morphologically "richer" than Chinese or "poorer" than a language like classical Hebrew, which marks verbs for both aspect and agreement with the subject (in *qaniti šloša s$^{\partial}$farim,* the verb *qaniti* is a first-person singular perfective form). Rather, such variation provides rich fodder for the study of morphological processing, insofar as speakers of different languages make different kinds of errors with morphology under demanding experimental conditions (Dick, Bates, & Ferstl, 2003) and present with different morphological impairments in the setting of brain damage (Bates, Friederici, & Wulfeck, 1987; Menn & Obler, 1990; Wulfeck, Bates, & Capasso, 1991). For example, Mandarin-speaking aphasic patients tend to omit functional morphemes (Packard, 1990), while Hebrew-speaking patients tend to make substitution errors with bound morphemes and omission errors with unbound functional morphemes (Friedmann & Grodzinksy, 1997).

Moreover, some theories posit the existence, in all languages, of morphemes that have no phonological content at all. In the phrase *two sheep,* for example, the plural marker is thought to be phonologically null. This is an exception for English, which generally marks plurals with the inflectional suffix-*s,* but is perhaps the rule for Mandarin, which has no marker for plurals per se (*shū* can mean either "book" or "books," depending on the context). In any given language, there may be very many grammatical features that are encoded by such zero morphemes (Pesetsky, 1995). On this view, nearly every word produced by a speaker or comprehended by a listener is, in fact, an agglomeration of lexical and functional morphemes, which convey various kinds of information crucial for encoding and decoding the meaning of that word in the context of an utterance. In other words,

KEVIN A. SHAPIRO Department of Psychology, Harvard University, Cambridge; Department of Medicine, Children's Hospital, Boston, Massachusetts

ALFONSO CARAMAZZA Department of Psychology, Harvard University, Cambridge, Massachusetts; Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy

morphology is not an optional stage in the production and comprehension of language, but an obligatory one (see also Shapiro, Shelton, & Caramazza, 2000).

Morphological composition in the lexicon

If many words appear to be composed of more than one morpheme, then must all morphologically complex words be built from scratch every time they are produced? Or are complex forms stored in the mental lexicon, so that they can be retrieved as prebuilt units when needed in speech or in comprehension?[1]

In some morphological domains, and particularly in the domain of inflectional morphology, it seems unlikely that all forms are fully listed in the lexicon (Butterworth, 1983). Inflected words have meanings that are transparent and predictable from their component parts, and regular inflections can be productively applied to novel words and nonwords that have no corresponding lexical representations (Berko Gleason, 1958; Goodglass & Berko, 1960). Furthermore, in some languages, the number of possible inflectional forms for any given word is so large as to render the full-listing hypothesis computationally infeasible. For example, in Finnish a noun may have as many as 2,000 distinct inflectional forms. Neuropsychological evidence also provides some support for the "online" nature of inflectional morphology: some patients appear to have selective problems with morphological inflection (Laine, Niemi, Koivuselkä-Sallinen, & Hyönä, 1995; Miceli & Caramazza, 1988), and others display different patterns of errors with lexical and inflectional morphemes (Miceli, Capasso, & Caramazza, 2004).

Some derivational processes are also likely to take place online, like the generation of novel derived forms with transparent meanings (*undigitizable*, *misunderestimate*) and productive lexical compounding, as in the Mandarin *wǒ mǎi shū le* (analogous to the English "I went book-buying") (Butterworth, 1983; Cutler, 1981; Vannest & Boland, 1999). For conventional derived words (like *conventional* and *composition*), however, the question is much more controversial (McQueen & Cutler, 1998). According to one set of theories, the lexicon has a root- (or stem-) based organization, meaning that the root is the basic unit that is stored in memory; derived forms must be produced by composition online (Forster, 1976; Taft, 1979; Taft & Forster, 1975). Other theories hold that common derived forms are stored as units, though they may be linked to other representations within the same derivational family (*compose*, *composure*, *composite*, *composition, decompose,* and so on (Caramazza, Miceli, Silveri, & Laudanna, 1985). A third kind of model proposes that only derived words that are highly frequent (as well as inflected forms that are highly frequent) are stored as units; low-frequency forms must be constructed or parsed online (Luzzatti, Mondini, & Semenza, 2001). The tendency to store derived forms in toto, as opposed to storing them in a decomposed manner, may vary across languages (Vannest, Bertram, Järvikivi, & Niemi, 2002).

The data available to distinguish between these cognitive models are equivocal, and there is very little evidence from cognitive neuroscience on the processing of morphologically derived words. Electrophysiological data suggest that morphological relationships between words affect lexical access at an earlier stage than phonological relationships (Pylkkänen, Feintuch, Hopkins, & Marantz, 2004), lending support to the general idea that morphological structure is represented in the lexicon but not distinguishing between the competing proposals about morphological decomposition. From this point forward we will generally be concerned with the cognitive neuroscience of productive morphological processes, and especially of morphological inflection.

Neural basis of productive morphological processes

Although morphological processing plays a central role in language, morphology has received comparatively little attention from cognitive neuroscientists. Much of what is known about morphology from a neuroscientific perspective comes from the study of brain-damaged patients who have difficulties in producing and understanding morphologically complex words. Impairment with the use of functional morphemes is classically observed in patients with lesions affecting the left inferior frontal lobe (Goodglass & Berko, 1960; Menn & Obler, 1990) and is considered a defining characteristic of agrammatic aphasia. By contrast, patients with posterior perisylvian lesions often show preservation of functional and derivational morphemes despite impairment in access to lexical roots. Such patients may produce apparently morphologically complex neologisms like *persessing* and *devorodation* (Buckingham & Kertesz, 1974; Caplan, Kellar, & Locke, 1972; Semenza, Butterworth, Panzeri, & Ferreri, 1990).

This generalization about anterior and posterior aphasics is not ironclad, however, as some patients with anterior lesions also appear to have preserved knowledge about functional morphology in the face of significant impairments in lexical access. For example, patient HG could produce *achooing* and *looming* but not *sneezing* or *weaving*, and she was also able to produce inflected nonwords in phrases like *he wugs* (Shapiro & Caramazza, 2003b). Patients with acquired "deep" dyslexia, which can arise as a sequela of heterogeneous and diffuse lesions in the left hemisphere, also characteristically make morphological errors in reading aloud (e.g., *initiative* for *initiate*) (Patterson, 1980).

On the one hand, the observation that morphological elements in language can be spared or impaired selectively in aphasia suggests that morphological rules rely on neural

circuits distinguishable from those that underly other aspects of language, like phonology and syntax. This suggestion coincides both with dominant linguistic theories and with psycholinguistic models of sentence processing, which postulate the independence of morphology from phonology and syntax on the basis of evidence like the differential involvement of lexical and functional morphemes in slips of the tongue (e.g., *it waits to pay*) (Garrett, 1980; Levelt, 1989; Schwartz, 1987).

On the other hand, it is not obvious that all morphological errors in aphasia are actually attributable to deficits in morphological knowledge as such. It has been argued that the morphological errors made in reading by acquired dyslexic patients are not morphological in origin at all, but rather are actually semantic errors or visual errors (Badecker & Caramazza, 1987; Castles, Coltheart, Savage, Bates, & Reid, 1996; Funnell, 1987; Plaut & Shallice, 1993). This proposal has been notoriously difficult to refute, although there is evidence to suggest that such patients do not systematically produce morphological forms that are more frequent or more imageable (Rastle, Tyler, & Marslen-Wilson, 2006). Similar doubts exist about the nature of morphological errors in naming, repetition, and spontaneous speech. In a study of repetition errors in 26 aphasic patients, Miceli and colleagues (2004) demonstrated that aphasic patients who make morphological errors invariably also make phonological errors, implying either that the neural circuits important for morphology are distinct but grossly inseparable from regions involved in phonological processing, or that what appear to be morphological errors are in fact errors of phonology.

In this case, there is some evidence to support both positions. Neuroimaging studies have found that the left inferior prefrontal cortex is recruited in a wide variety of linguistic and nonlinguistic tasks, including the processing of grammatical gender (Miceli et al., 2002), phonological processing (Heim & Friederici, 2003; Heim, Opitz, Muller, & Friederici, 2003; Indefrey & Levelt, 2000), and phonological working memory (Hickok & Poeppel, 2007; Paulesu, Frith, & Frackowiak, 1993). These diverse results suggest that the left anterior perisylvian region may contain populations of neurons that are heterogeneous in function. It may be that any brain lesion that is large enough or severe enough to disrupt morphological processes will also disrupt phonological processes—and perhaps other cognitive functions as well. Alternatively, perhaps it is the case that functional morphemes are especially vulnerable to impairment because of the extra demands they place on the phonemic processor. Kean observed that agrammatic patients tend to omit functional morphemes that are phonologically less salient (Kean, 1978, 1979). Both children and some aphasic patients fare more poorly with inflections that are phonologically more complex (Berko Gleason, 1958; Shapiro & Caramazza, 2003a; Bird, Lambon-Ralph, Seidenberg, McClelland, & Patterson, 2003; Joanisse & Seidenberg, 1999; Patterson, Lambon-Ralph, Hodges, & McClelland, 2001).

An intermediate possibility is that morphology is neither an autonomous function nor a wholly owned subsidiary of phonology, but rather a confederation of processes that operate at the interstices of language, ensuring that abstract syntactic representations with particular grammatical and structural features can be matched with specific, contextually appropriate lexical representations, and that these lexical representations in turn can be converted into phonological strings. Morphological deficits in aphasia may arise when one of these interfaces is compromised by damage at a particular level of language processing. For instance, patients who have difficulty with lexical access may be prone to making paragrammatic substitution errors in either functional or derivational morphology (Caplan et al., 1972; Kohn & Melvold, 1999; Laine et al., 1995; Miceli & Caramazza, 1988; Semenza et al., 1990). Such patients may make relatively few phonological errors, especially if their errors in other language production tasks are not primarily phonological—as was true for patient HH described by Laine and colleagues, who produced paralexias involving both functional morphemes and root (or stem) morphemes (e.g., *pesä+lla* "on the base" was read as *maila+sta* "from the bat"). Interestingly, this patient's lesion largely spared the left inferior prefrontal cortex, but may have involved subcortical connections between the left frontal lobe and posterior perisylvian areas that were also damaged.

Some patients appear simply to ignore morphemes that are not lexical roots, even when access to phonological information appears to be intact (Tyler, Behrens, Cobb, & Marslen-Wilson, 1990); in these cases, the deficit may also occur at the level of lexical retrieval. However, patients with postlexical processing deficits may have particular difficulty with functional morphemes, which are often unstressed and can require the resyllabification of words and phrases (Kean, 1978). Likewise in comprehension, patients of this type may have difficulty parsing functional affixes (Tyler & Cobb, 1987). Others still may present with morphological impairments that are linked to the ability to use particular kinds of syntactic information (Goodglass & Berko, 1960), such as information about tense (Friedmann & Grodzinksy, 1997; Miceli, Silveri, Romani, & Caramazza, 1989) or knowledge about a specific grammatical category (Laiacona & Caramazza, 2004; Shapiro & Caramazza, 2003a; Tsapkini, Jarema, & Kehayia, 2002).

The question of how morphology interacts with other subcomponents of language, as well as with domain-general mechanisms in cognitive processing, has proven to be a fruitful field of research in this otherwise relatively uncultivated domain of cognitive neuroscience. We will discuss two examples in the sections that follow. First, there is the

controversy over the representation of phonologically regular and irregular morphological forms (like *ducks* and *geese,* respectively), which addresses a question at the interface of morphology and phonology: namely, how are abstract morphological features converted to phonological information? Second, there is the debate about whether the brain distinguishes between categories of words, like nouns and verbs, by virtue of their abstract morphosyntactic properties. Both topics demonstrate that the neuroscientific investigation of morphological processing can generate findings and novel hypotheses that are relevant not only to the biology of language, but also to the understanding of cognition more broadly.

The regular/irregular debate

As we saw with the example *two sheep,* a morphologically marked grammatical feature (like past tense) may be expressed phonologically in more than one way within a language. When this is the case, there is usually only a small set of regular morphophonological transformations that can be applied productively to novel words (e.g., *one google/many googles*); the other transformations tend to be frozen and nonproductive (e.g., *goose/geese* but not *google/*geegle*). This discrepancy has led many to propose that regular and irregular forms are processed by separate cognitive and, by extension, neurobiological mechanisms (Pinker, 1991).

At first glance, this proposal appears to be supported by the existence among aphasic patients of double dissociations in processing regular and irregular morphology (Marslen-Wilson & Tyler, 1997; Tyler, deMornay-Davies, et al., 2002; Tyler, Randall, & Marslen-Wilson, 2002; Ullman et al., 1997); see figure 53.1. The evidence for such dissociations has been challenged, however, on the grounds that they are reducible to phonological factors (Bird et al., 2003; Braber, Patterson, Ellis, & Ralph, 2005; Lambon-Ralph, Braber, McClelland, & Patterson, 2005) or to a combination of phonological and semantic deficits (Joanisse & Seidenberg, 1999; Patterson et al., 2001): in other words, regular words are more susceptible to errors because they involve transformations that are phonologically more complex, not because they are processed by a different route than regular morphological transformations. Coincidentally, it has been reported that patients with impairments in regular morphology tend to have lesions in left inferior prefrontal regions (Ullman et al., 1997)—a part of the brain which, we have seen, is also thought to be important for phonological processing.

Whether the single-route account can successfully explain all cases of regular/irregular dissociations is a matter of some debate, and there appear to be at least some anterior aphasic patients for whom a selective deficit in irregular morphology co-occurs with a deficit in phonological processing, contrary to what this account would seem to predict (Miozzo, 2003). The dual-route account, by contrast, is often (albeit not of necessity) linked with a different hypothesis about the functional role of the left inferior frontal cortex. Ullman and colleagues have proposed that the processing of regular forms depends on a frontal-striatal circuit that is engaged more generally in the representation of procedural memory, not necessarily limited to language (Pinker & Ullman, 2002; Ullman et al., 1997; Ullman, 2001, 2004; Ullman et al., 2005). However, irregular forms are said to depend on temporal lobe structures that are important for the representation of declarative memory.

The declarative/procedural hypothesis is attractive in that it situates a linguistic dissociation within the context of a general model in cognitive neuroscience. However, the empirical data have so far been mixed. To begin with, the putative link between frontal lobe damage and impairments in regular morphological processing is tenuous at best. Although some patients with anterior lesions have selective deficits for regular morphological transformations, and some patients with temporal lobe damage have more difficulty with irregular transformations (Tyler, deMornay-Davies, et al., 2002; Tyler, Randall, et al., 2002; Ullman et al., 1997), there are numerous counterexamples of patients with frontal lobe damage who have greater difficulty with irregular words (de Diego Balaguer, Costa, Gallés, Juncadella, & Caramazza, 2004; Penke, Janssen, & Krause, 1999; Shapiro & Caramazza, 2003a). Neuroimaging studies of regular and irregular inflection have not clarified the issue: some have failed to elucidate distinctions in the cortical regions activated by regular and irregular stimuli (Sahin, Pinker, & Halgren, 2006), others have shown that irregular words elicit greater activity in prefrontal regions (de Diego Balaguer et al., 2006; Desai, Conant, Waldron, & Binder, 2006), and still others seem to indicate that regular words produce greater activation in both frontal and temporal regions (Beretta et al., 2003; Tyler, Stamatakis, Post, Randall, & Marslen-Wilson, 2005).

This confusion may be due in part to the fact that not all regular/irregular dissociations arise at the same level of language processing. As Druks observes, the regular/irregular distinction is relevant only in the domain of morphophonology, and so the dissociation predicted by the procedural/declarative hypothesis should be evident only when rule-based morphophonological processing is emphasized (Druks, 2006). A corollary is that deficits at other levels may also interact with the processing of regular and irregular word forms. For at least one of the patients who exhibited a dissociation not in line with the procedural/declarative model, the locus of impairment appeared to be not morphophonological, but morphosyntactic: the patient produced regular forms better than irregular forms, but only when the stimuli were inflected verbs (Shapiro & Caramazza, 2003a). This result reinforces the observation that the left inferior

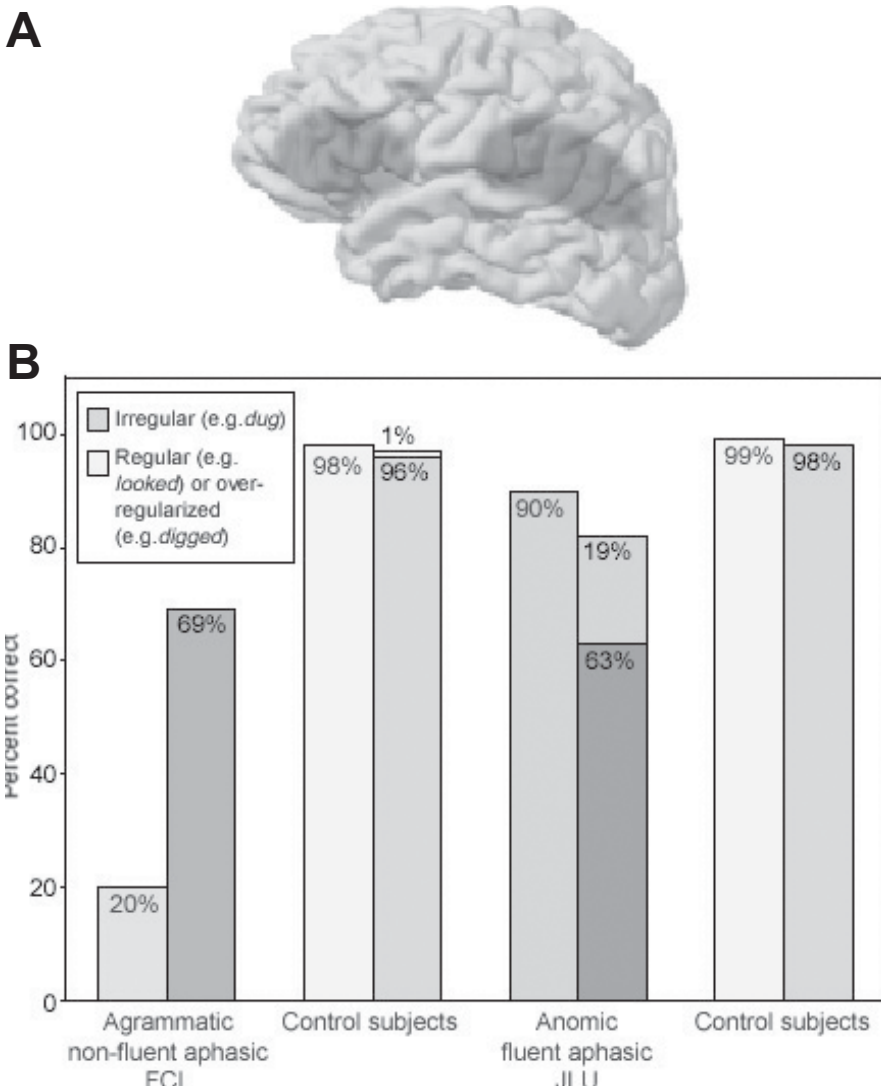


FIGURE 53.1 A neuropsychological dissociation in processing regular and irregular verb forms. (*A*) The approximate lesion sites of patient FCL (red area, left anterior perisylvian regions), who had symptoms of agrammatism, and patient JLU (green area, left temporoparietal region), who had symptoms of anomia. (*B*) Results of verb inflection tests showed that the agrammatic patient had more trouble inflecting regular verbs (lighter bars) than irregular verbs (darker bars), whereas the anomic patient had more trouble inflecting irregular verbs—and overapplied the regular suffix to many of the irregulars (light green bar on top of dark green bar). The performance of age- and education-matched control subjects is shown in the gray bars. (Reprinted from Pinker & Ullman, 2002.) (See color plate 65.)

prefrontal cortex is functionally heterogeneous. Not all patients with lesions in this area may be expected to have the same pattern of linguistic performance, and activation of this region in neuroimaging paradigms may be particularly sensitive to the demands of the task that is employed.

A second anatomical claim of the procedural/declarative hypothesis is that the basal ganglia, and specifically the striate nuclei (the caudate and putamen), are crucial for the processing of grammatical rules. Studies of patients with early Huntington's disease, which first affects the caudate, have shown that these patients are indeed impaired in producing morphologically complex word forms (Gordon & Illes, 1987) and making rule-based judgments about such forms (Teichmann, Dupoux, Kouider, & Bachoud-Lévi, 2006). They may make more errors than control subjects in the production of regularly inflected words (Longworth, Keenan, Barker, Marslen-Wilson, & Tyler, 2005), though the latter finding appears to be subtle and task dependent. Some neuroimaging experiments corroborate the idea that the caudate nuclei are particularly active in the detection of syntactic anomalies, including anomalies signaled by morphological structure (Forkstam, Hagoort, Fernandez, Ingvar, & Petersson, 2006; Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004; Moro et al., 2001). The evidence therefore seems, on balance, to support a role for the caudate in the application of linguistic rules. By contrast, patients with

nonstriatal basal ganglia lesions, like those with Parkinson's disease, do not reliably display either regular/irregular effects (Longworth et al.) or difficulties with syntactic rules (Reber & Squire, 1999; Small, Lyons, & Kemper, 1997; Witt, Nühsman, & Deuschl, 2002).

Grammatical categories and morphological processes

It has long been observed that some patients with aphasia have more difficulty producing nouns than verbs, while others show the opposite dissociation (Goodglass, Klein, Carey, & Jones, 1966; Luria & Tsvetkova, 1967; Miceli, Silveri, Villa, & Caramazza, 1984). This dissociation is particularly relevant to the discussion of morphological knowledge, since nouns and verbs can be distinguished formally by virtue of the fact that they undergo different kinds of morphological operations. For example, nouns in English are marked for number and possession, while verbs are marked for tense and subject agreement. The differential distribution of morphological operations over nouns and verbs may also facilitate the acquisition of grammatical categories in the course of language development (Maratsos & Chalkley, 1981).

In some cases, "grammatical" category deficits may not be grammatical in origin at all; instead, they may be attributable to impairments in access to specific kinds of semantic information associated more with words of one category than another. For example, many verbs refer to actions; a selective deficit in verb production could, therefore, result from damage to brain regions involved in representing knowledge about actions. Likewise, an apparent noun deficit could result from an underlying impairment in retrieving semantic knowledge about concrete objects (Damasio & Tranel, 1993; McCarthy & Warrington, 1985).

Still, some patients have deficits that cannot readily be explained by positing a breakdown in access to semantic knowledge. A number of patients with deficits in noun or verb production are particularly impaired in producing morphologically inflected words when they clearly belong to one grammatical category (Laiacona & Caramazza, 2004; Shapiro & Caramazza, 2003a; Shapiro et al., 2000; Tsapkini et al., 2002), even when, in some cases, they are able to produce exactly the same strings when the context indicates that they belong to the other category (Shapiro & Caramazza, 2003a; Shapiro et al.). For example, patient RC, described by Shapiro & Caramazza (2003a), was able to produce the phrase *these judges*, but not the phrase *he judges*. This deficit extended even to the production of morphologically inflected neologisms (e.g., *these wugs* versus *he wugs*).

The fact that these patients also have difficulties with the affected category in production tasks that do not explicitly involve morphological inflection—like picture naming and delayed repetition—cannot be taken as prima facie evidence against a morphological locus for their deficits; to the contrary, it reinforces the idea that morphological processing is an essential step in all lexical output (Shapiro et al., 2000). Moreover, a category-specific impairment in morphological inflection is not the simple consequence of any deficit that appears to affect noun or verb production: some patients have difficulty producing words of one category without any evident impairment in morphology (Shapiro & Caramazza, 2003b).

What brain regions subserve the morphological processing of nouns and verbs? The data on this score are limited, as the patients who have clear grammatical impairments are few in number and have lesions that are too extensive and ill-defined to allow for meaningful comparisons. However, it has been shown that when neurologically intact subjects perform morphological transformation tasks following transcranial magnetic stimulation, a processing deficit for verbs emerges only when the stimulation is targeted to a focal area of the left middle frontal gyrus, situated near the triangular portion of Broca's area (Cappelletti, Fregni, Shapiro, Pascual-Leone, & Caramazza, 2008; Shapiro, Pascual-Leone, Mottaghy, Gangitano, & Caramazza, 2001); see figure 53.2. This area is included within the lesion of patient RC (Shapiro & Caramazza, 2003a).

Curiously, neuroimaging studies have offered little, if any, support for the notion that this part of the brain distinguishes nouns from verbs on the basis of morphology. In most previous studies, differences in activation between nouns and verbs have been observed not in the middle frontal gyrus but in the left inferior frontal gyrus, or Broca's area (Longe, Randall, Stamatakis, & Tyler, 2007; Perani et al., 1999; Sahin et al., 2006; Tyler, Bright, Fletcher, & Stamatakis, 2004). Although this area is also included within the lesions of patients with apparent verb deficits, in most cases the lesions are much more extensive than the area of activation (figure 53.3). Moreover, Tyler and colleagues (Longe et al., 2007; Tyler et al.) have argued that the increased activation they observed for verbs in this region does not reflect categorical specificity as such, but rather, the greater complexity of verb morphology as compared to noun morphology. For example, they point out that verb inflection in English may be bound up with the computation of long-distance syntactic dependencies, like agreement with the subject of a sentence. It is indeed likely that the inferior frontal gyrus is sensitive to any variation in processing demands, but this may be more a property of particular stimuli than of the English language in general: nouns with low-frequency and atypical inflectional forms activate this area more than do verbs (Sahin et al.). When noun and verb stimuli are carefully matched for difficulty, targeted suppression of Broca's area with transcranial magnetic stimulation results in a category-general delay in processing, but no category-specific effects (Cappelletti et al., 2008).

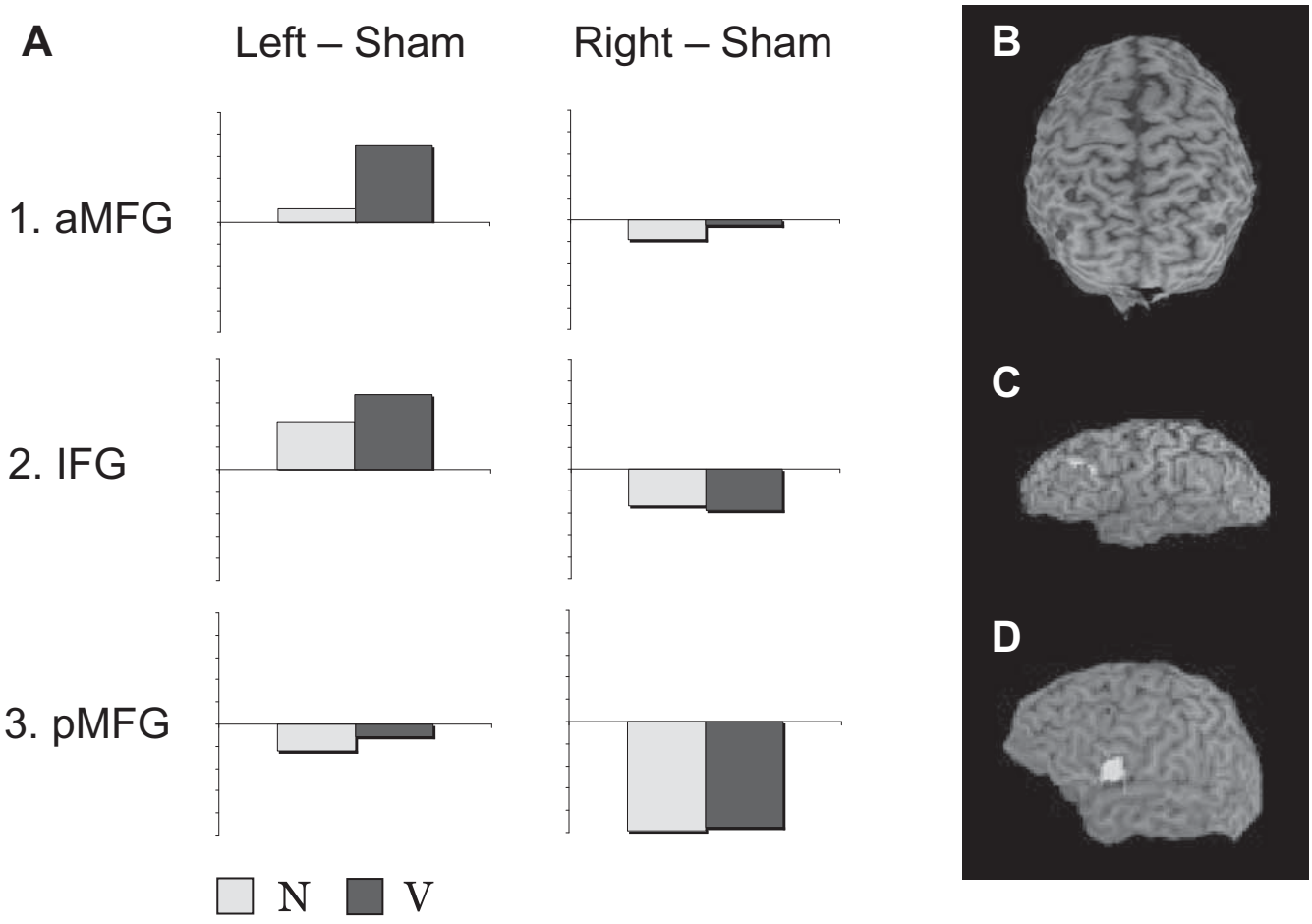


FIGURE 53.2 Results of the rTMS experiment reported by Cappelletti et al. (2008), showing a selective disruption in verb processing following stimulation to the left anterior frontal gyrus. (*A*) The mean difference in reaction times to nouns and verbs with repetitive TMS compared to sham stimulation in three areas: the anterior middle frontal gyrus (aMFG), inferior frontal gyrus (IFG), and posterior middle frontal gyrus (pMFG). (*B*) The sites of stimulation to the IFG and pMFG. The remaining panels demonstrate the stereotactic application of TMS to the left pMFG (*C*) and left IFG (*D*). (Modified from Cappelletti, Fregni, Shapiro, Pascual-Leone, & Caramazza, 2008.) (See color plate 66.)

We propose that the left inferior frontal gyrus represents a common pathway for the production of words bearing functional morphemes that specify grammatical information relevant to one category or another. In other words, this area (perhaps along with the striate nuclei of the basal ganglia) may be important for the conversion of morphological elements into phonological segments. The process of selecting syntactically appropriate functional morphemes may be handled by different upstream regions, like the left anterior middle frontal gyrus for verbs. These morphosyntactic regions, in turn, must normally receive information from the lexicon, with the constraint that only words meeting certain requirements should be processed as nouns or verbs—allowing us to say, for example, that *he rose to smell* or *he smelled a rose*, but not *he has been rosing up the place all afternoon.*

The hypothesis of a neuroanatomical dissociation between grammatically based morphological processes and form-based morphological processes also has the virtue of accounting for certain striking phenomena that have hitherto been somewhat difficult to reconcile with other theories about the organization of language in the brain: namely, the finding that some aphasic patients exhibit grammatical category deficits that are restricted to either spoken or written output (Caramazza & Hillis, 1991; Hillis & Caramazza, 1995; Hillis, Tuffiash, & Caramazza, 2002; Hillis, Wityk, Barker, & Caramazza, 2003; Rapp & Caramazza, 2003). Perhaps the clearest example of this kind of modality specific deficit is the case of patient KSR, who produced verbs better than nouns in speech, but nouns better than verbs in writing (Rapp & Caramazza, 2002). That such patients are able to produce the same stimuli in one modality but not in another strongly implies that the patients' problems do not arise at the semantic level of representation. Instead, it has been proposed that the cortical regions responsible for storing and accessing lexical representations are segregated along lines of both modality and grammatical category, so that brain

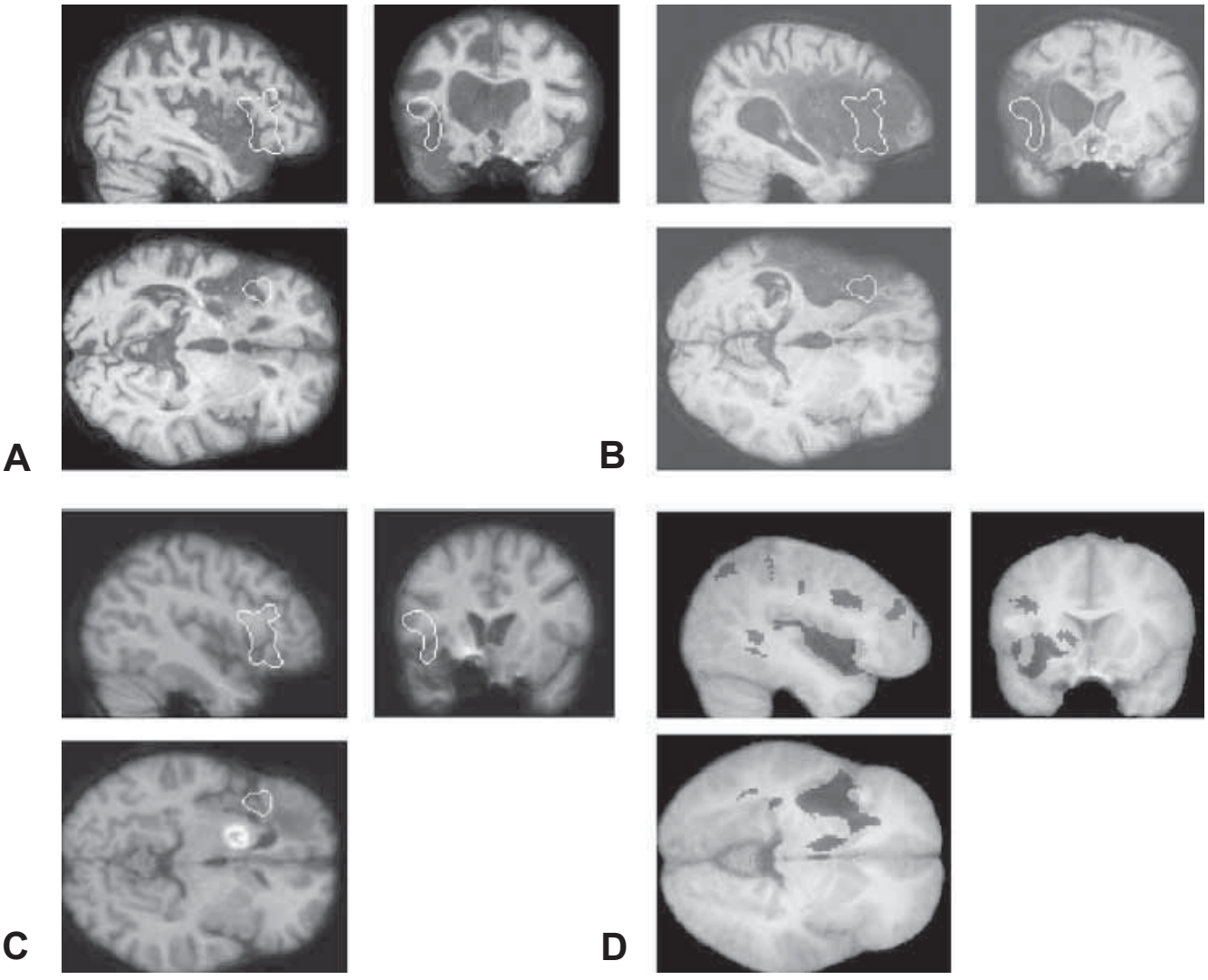


FIGURE 53.3 The area found by Tyler and colleagues (2004) to be more active for inflected verbs than inflected nouns in an fMRI semantic judgment paradigm, compared to the lesion sites of three aphasic patients with deficits in processing regularly inflected verb forms in a priming task. (*A–C*) T1-weighted MR images of three patients with an outline of the activation found in the verbs-nouns contrast superimposed on them. (*D*) A mean of the spatially normalized T1 images of the 12 subjects in the fMRI experiment overlaid with the lesion overlap of the three patients in *A–C*. Lesion overlap is shown in blue, the significant activation found in the verbs-nouns contrast is in yellow, and the overlap between common lesion volume of the three patients and the activation is in green. (Reprinted from Tyler, Bright, Fletcher, & Stamatakis, 2004.) (See color plate 67.)

damage might selectively affect access to orthographic verb representations, for example.

While this proposal is not logically impossible, it is somewhat difficult to reconcile with the fact that these patients' lesions tend to be relatively large, and the areas implicated—like the left posterior inferior frontal and precentral gyri in two patients unable to write verbs (Hillis et al., 2003)—are unlikely candidates for modality-specific lexical stores. However, if we suppose that morphosegmental processes (in phonology and orthography) are dissociable from lexical retrieval and morphosyntactic feature selection, an alternative explanation becomes available. It may be that modality-specific grammatical-class deficits are manifestations of disconnections between morphosyntactic processors, segregated by grammatical category, and morphosegmental processors, which may be segregated by modality.

Precisely what brain areas are important for category-specific morphosyntactic processes and for the representation of phonological and orthographic segments is, of course, a matter that requires much further investigation. With respect to morphosyntactic processing, the rTMS studies reviewed here suggest that the anterior portion of the left middle frontal gyrus may be crucial for verbs (Cappelletti et al., 2008; Shapiro et al., 2001). The data for nouns are even more severely limited: the lesion data implicate either the left inferior frontal lobe or the inferior parietal lobe

(Shapiro et al., 2000), although none of the frontal areas tested with rTMS was found to be crucial for nouns. What is clear, however, is that some components of the neural circuitry for language production are sensitive to information about grammatical category, while others are dedicated to the processing of particular kinds of output. We believe that this hypothesized division of labor, with the ultimate goal of constructing morphemes into producible and comprehensible words, may provide a productive framework for investigating the neurobiological mechanisms by which language operates.

NOTE

1. In this chapter we are concerned primarily with language production: in other words, how do speakers produce morphologically complex words? Of course, an analogous problem exists in the domain of comprehension: how do listeners access the meaning of morphologically complex words? We make the assumption here that the lexicon is unitary—that is, that the same kinds of lexical representations are accessed in production and comprehension. It follows that theories about morphological composition in the lexicon, even those based empirically on evidence from comprehension tasks, should also apply to language production.

REFERENCES

Badecker, W., & Caramazza, A. (1987). The analysis of morphological errors in a case of acquired dyslexia. *Brain Lang., 32*, 278–305.

Bates, E., Friederici, A., & Wulfeck, B. (1987). Grammatical morphology in aphasia: Evidence from three languages. *Cortex, 23*(4), 545–574.

Beretta, A., Campbell, C., Carr, T. H., Huang, J., Schmitt, L. M., Christianson, K., et al. (2003). An ER-fMRI investigation of morphological inflection in German reveals that the brain makes a distinction between regular and irregular forms. *Brain Lang., 85*, 67–92.

Berko Gleason, J. (1958). The child's learning of English morphology. *Word, 14*, 150–177.

Bird, H., Lambon-Ralph, M. A., Seidenberg, M. S., McClelland, J. L., & Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? *J. Mem. Lang., 48*, 502–526.

Braber, N., Patterson, K., Ellis, K., & Ralph, M. A. L. (2005). The relationship between phonological and morphological deficits in Broca's aphasia: Further evidence from errors in verb inflection. *Brain Lang., 92*(3), 278–287.

Buckingham, H. W., & Kertesz, A. (1974). A linguistic analysis of fluent aphasia. *Brain Lang., 1*(1), 43–61.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production* (Vol. 2). London: Academic Press.

Caplan, D., Kellar, L., & Locke, S. (1972). Inflection of neologisms in aphasia. *Brain, 95*(1), 169–172.

Cappelletti, M., Fregni, F., Shapiro, K., Pascual-Leone, A., & Caramazza, A. (2008). Processing nouns and verbs in the left frontal cortex: A TMS study. *J. Cogn. Neurosci., 20*(4), 707–720.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature, 349*, 788–790.

Caramazza, A., Miceli, G., Silveri, M. C., & Laudanna, A. (1985). Reading mechanisms and the organization of the lexicon: Evidence from acquired dyslexia. *Cogn. Neuropsychol., 2*, 81–114.

Castles, A., Coltheart, M., Savage, G., Bates, A., & Reid, L. (1996). Morphological processing and visual word recognition: Evidence from acquired dyslexia. *Cogn. Neuropsychol., 13*, 1041–1057.

Cutler, A. (1981). Degrees of transparency in word formation. *Can. J. Ling., 26*, 73–77.

Damasio, A. R., & Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proc. Natl. Acad. Sci. USA, 90*(11), 4957–4960.

De Diego Balaguer, R., Costa, A., Gallés, N. S., Juncadella, M., & Caramazza, A. (2004). Regular and irregular morphology and its relation with agrammatism: Evidence from Spanish and Catalan. *Cortex, 40*(1), 157–158.

De Diego Balaguer, R., Rodríguez-Fornells, A., Rotte, M., Bahlmann, J., Heinze, H.-J., & Münte, T. F. (2006). Neural circuits subserving the retrieval of stems and grammatical features in regular and irregular verbs. *Hum. Brain Mapp., 27*, 874–888.

Desai, R., Conant, L. L., Waldron, E., & Binder, J. R. (2006). fMRI of past tense processing: The effects of phonological complexity and task difficulty. *J. Cogn. Neurosci., 18*(2), 278–297.

Dick, F., Bates, E., & Ferstl, E. C. (2003). Spectral and temporal degradation of speech as a simulation of morphosyntactic deficits in English and German. *Brain Lang., 85*(3), 535–542.

Druks, J. (2006). Morpho-syntactic and morpho-phonological deficits in the production of regularly and irregularly inflected verbs. *Aphasiology, 20*(9), 993–1017.

Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M., & Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *NeuroImage, 32*(2), 956–967.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North Holland.

Friedmann, N. A., & Grodzinksy, Y. (1997). Tense and agreement in agrammatic production: Pruning the syntactic tree. *Brain Lang., 56*(3), 397–425.

Funnell, E. (1987). Morphological errors in acquired dyslexia: A case of mistaken identity. *Q. J. Exp. Psychol. [A], 39*, 497–539.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177–220). New York: Academic Press.

Goodglass, H., & Berko, J. (1960). Agrammatism and inflectional morphology in English. *J. Speech Hear. Res., 3*, 257–267.

Goodglass, H., Klein, B., Carey, P., & Jones, K. (1966). Specific semantic word categories in aphasia. *Cortex, 2*(1), 74–89.

Gordon, W. P., & Illes, J. (1987). Neurolinguistic characteristics of language production in Huntington's disease: A preliminary report. *Brain Lang., 31*(1), 1–10.

Heim, S., & Friederici, A. D. (2003). Phonological processing in language production: Time course of brain activity. *NeuroReport, 14*(16), 2031–2033.

Heim, S., Opitz, B., Muller, K., & Friederici, A. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Brain Res. Cogn. Brain Res., 16*, 285–296.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci., 8*(5), 393–402.

Hillis, A. E., & Caramazza, A. (1995). Representation of grammatical knowledge in the brain. *J. Cogn. Neurosci., 7*, 396–407.

Hillis, A. E., Tuffiash, E., & Caramazza, A. (2002). Modality-specific deterioration in naming verbs in nonfluent primary progressive aphasia. *J. Cogn. Neurosci., 14*(7), 1099–1108.

Hillis, A. E., Wityk, R. J., Barker, P. B., & Caramazza, A. (2003). Neural regions essential for writing verbs. *Nat. Neurosci., 6*(1), 19–20.

Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 845–865). Cambridge, MA: MIT Press.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury. *Proc. Natl. Acad. Sci. USA, 96*, 7592–7597.

Kean, M.-L. (1978). The linguistic interpretation of aphasic syndromes: Agrammatism in Broca's aphasia, an example. *Cognition, 5*, 9–46.

Kean, M.-L. (1979). Agrammatism: A phonological deficit? *Cognition, 7*, 69–84.

Kohn, S. E., & Melvold, J. (1999). Effects of morphological complexity on phonological output deficits. *Brain Cogn. 40*(1), 156–159.

Laiacona, M., & Caramazza, A. (2004). The noun/verb dissociation in language production: Varieties of causes. *Cogn. Neuropsychol., 21*, 103–124.

Laine, M., Niemi, J., Koivuselkä-Sallinen, P., & Hyönä, J. (1995). Morphological processing of polymorphemic nouns in a highly inflecting language. *Cogn. Neuropsychol., 12*(5), 457–502.

Lambon-Ralph, M. A., Braber, N., McClelland, J. L., & Patterson, K. (2005). What underlies the neuropsychological pattern of irregular > regular past tense verb production? *Brain Lang., 93*, 106–119.

Levelt, W. J. M. (1989). *Speaking.* Cambridge, MA: MIT Press.

Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *J. Cogn. Neurosci., 16*(3), 427–438.

Longe, O., Randall, B., Stamatakis, E. A., & Tyler, L. K. (2007). Grammatical categories in the brain: The role of morphological structure. *Cereb. Cortex, 17*(8), 1812–1820.

Longworth, C. E., Keenan, S. E., Barker, R. A., Marslen-Wilson, W. D., & Tyler, L. K. (2005). The basal ganglia and rule-governed language use: Evidence from vascular and degenerative conditions. *Brain, 2005*(128), 584–596.

Luria, A. R., & Tsvetkova, L. S. (1967). Towards the mechanisms of "dynamic aphasia." *Acta Neurol. Psychiatr. Belg., 67*(11), 1045–1057.

Luzzatti, C., Mondini, S., & Semenza, C. (2001). Lexical representation and processing of morphologically complex words: Evidence from the reading performance of an Italian agrammatic patient. *Brain Lang., 79*(3), 345–359.

Maratsos, M., & Chalkley, M. A. (1981). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), *Children's language* (pp. 127–214). New York: Gardner Press.

Marslen-Wilson, W., & Tyler, L. K. (1997). Dissociating types of mental computation. *Nature, 387*(6633), 592–594.

McCarthy, R. A., & Warrington, E. K. (1985). Category specificity in an agrammatic patient: The relative impairment of verb retrieval and comprehension. *Neuropsychologia, 23*(6), 709–727.

McQueen, J. M., & Cutler, A. (1998). Morphology in word recognition. In A. Spencer & A. M. Zwicky (Eds.), *Handbook of morphology* (pp. 406–427). Oxford, UK: Blackwell.

Menn, L., & Obler, L. K. (1990). *Agrammatic aphasia: A cross-language narrative sourcebook.* Philadelphia: John Benjamins.

Miceli, G., Capasso, R., & Caramazza, A. (2004). The relationships between morphological and phonological errors in aphasic speech: Data from a word repetition task. *Neuropsychologia, 42*(3), 273–287.

Miceli, G., & Caramazza, A. (1988). Dissociation of inflectional and derivational morphology. *Brain Lang., 35*(1), 24–65.

Miceli, G., Silveri, M. C., Romani, C., & Caramazza, A. (1989). Variation in the pattern of omissions and substitutions of grammatical morphemes in the speech of so-called agrammatic patients. *Brain Lang., 36*, 447–492.

Miceli, G., Silveri, M. C., Villa, G., & Caramazza, A. (1984). On the basis for the agrammatic's difficulty in producing main verbs. *Cortex, 20*(2), 207–220.

Miceli, G., Turriziani, P., Caltagirone, C., Capasso, R., Tomaiuolo, F., & Caramazza, A. (2002). The neural correlates of grammatical gender: An fMRI investigation. *J. Cogn. Neurosci., 14*(4), 618–628.

Miozzo, M. (2003). On the processing of regular and irregular forms of verbs and nouns. *Cognition, 87*, 101–127.

Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S. F., & Fazio, F. (2001). Syntax and the brain: Disentangling grammar by selective anomalies. *NeuroImage, 13*(1), 110–118.

Packard, J. L. (1990). Agrammatism in Chinese: A case study. In L. Menn & L. K. Obler (Eds.), *Agrammatic aphasia: A cross-language narrative sourcebook* (pp. 1191–1224). Philadelphia: John Benjamins.

Patterson, K. (1980). Derivational errors. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 286–306). London: Routledge & Kegan Paul.

Patterson, K., Lambon-Ralph, M. A., Hodges, J. R., & McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia, 39*, 709–724.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of verbal working memory. *Nature, 362*, 342–345.

Penke, M., Janssen, U., & Krause, M. (1999). The representation of inflectional morphology: Evidence from Broca's aphasia. *Brain Lang., 68*(1), 225–232.

Perani, D., Cappa, S. F., Schnur, T., Tettamanti, M., Collina, S., Rosa, M. M., & Fazio, F. (1999). The neural correlates of noun and verb processing: A PET study. *Brain, 122*(12), 2337–2344.

Pesetsky, D. (1995). *Zero morphology.* Cambridge, MA: MIT Press.

Pinker, S. (1991). Rules of language. *Science, 253*, 530–535.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends Cogn. Sci., 6*, 456–463.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia—A case-study of connectionist neuropsychology. *Cogn. Neuropsychol., 10*, 377–500.

Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: An MEG study. *Cognition, 91*(3), B35-B45.

Rapp, B., & Caramazza, A. (2002). Selective difficulties with spoken nouns and written verbs: A single case study. *J. Neurolinguistics, 15*(3–5), 373–402.

Rapp, B., & Caramazza, A. (2003). Selective difficulties with spoken nouns and written verbs: A single case study. *J. Neurolinguistics, 15*(3–5), 373–402.

Rastle, K., Tyler, L. K., & Marslen-Wilson, W. (2006). New evidence for morphological errors in deep dyslexia. *Brain Lang., 97*, 189–199.

Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behav. Neurosci., 113*(2), 235–242.

Sahin, N. T., Pinker, S., & Halgren, E. (2006). Abstract grammatical processing of nouns and verbs in Broca's area: Evidence from fMRI. *Cortex, 42*, 540–562.

Schwartz, M. F. (1987). Patterns of speech production deficit within and across aphasia syndromes: Application of a psycholinguistic model. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 163–199). Hove, Sussex, UK: LEA.

Semenza, C., Butterworth, B., Panzeri, M., & Ferreri, T. (1990). Word formation: New evidence from aphasia. *Neuropsychologia, 28*(5), 499–502.

Shapiro, K., & Caramazza, A. (2003a). Grammatical processing of nouns and verbs in left frontal cortex? *Neuropsychologia, 41*(9), 1189–1198.

Shapiro, K., & Caramazza, A. (2003b). Looming a loom: Evidence for independent access to grammatical and phonological properties in verb retrieval. *J. Neurolinguistics, 16*(2–3), 85–111.

Shapiro, K. A., Pascual-Leone, A., Mottaghy, F. M., Gangitano, M., & Caramazza, A. (2001). Grammatical distinctions in the left frontal cortex. *J. Cogn. Neurosci., 13*(6), 713–720.

Shapiro, K., Shelton, J., & Caramazza, A. (2000). Grammatical class in lexical production and morphological processing: Evidence from a case of fluent aphasia. *Cogn. Neuropsychol., 17*, 665–682.

Small, J. A., Lyons, K., & Kemper, S. (1997). Grammatical abilities in Parkinson's disease: Evidence from written sentences. *Neuropsychologia, 35*(12), 1571–1576.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Mem. Cogn., 7*, 263–272.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *J. Verb. Learn. Verb. Beh., 14*, 638–647.

Teichmann, M., Dupoux, E., Kouider, S., & Bachoud-Lévi, A. C. (2006). The role of the striatum in processing language rules: Evidence from word perception in Huntington's disease. *J. Cogn. Neurosci., 18*(9), 1555–1569.

Tsapkini, K., Jarema, G., & Kehayia, E. (2002). A morphological processing deficit in verbs but not in nouns: A case study in a highly inflected language. *J. Neurolinguistics, 15*(3), 265–288.

Tyler, L. K., Behrens, S., Cobb, H., & Marslen-Wilson, W. (1990). Processing distinctions between stems and affixes: Evidence from a non-fluent aphasic patient. *Cognition, 36*, 129–153.

Tyler, L. K., Bright, P., Fletcher, P., & Stamatakis, E. A. (2004). Neural processing of nouns and verbs: The role of inflectional morphology. *Neuropsychologia, 42*(4), 512–523.

Tyler, L. K., & Cobb, H. (1987). Processing bound grammatical morphemes in context: The case of an aphasic patient. *Lang. Cogn. Process., 2*(3), 245–262.

Tyler, L. K., deMornay-Davies, P., Anokhina, R., Longworth, C., Randall, B., & Marslen-Wilson, W. D. (2002). Dissociations in processing past tense morphology: Neuropathology and behavioral studies. *J. Cogn. Neurosci., 14*(1), 79–94.

Tyler, L. K., Randall, B., & Marslen-Wilson, W. D. (2002). Phonology and neuropsychology of the English past tense. *Neuropsychologia, 40*(8), 1154–1166.

Tyler, L. K., Stamatakis, E. A., Post, B., Randall, B., & Marslen-Wilson, W. (2005). Temporal and frontal systems in speech comprehension: An fMRI study of past tense processing. *Neuropsychologia, 43*(13), 1963–1974.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *J. Psycholinguist. Res., 30*, 37–69.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270.

Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growdon, J., Koroshetz, W., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are part of the procedural system. *J. Cogn. Neurosci., 9*, 266–276.

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain Lang., 93*(2), 185–238.

Vannest, J., Bertram, R., Järvikivi, J., & Niemi, J. (2002). Counterintuitive cross-linguistic differences: More morphological computation in English than in Finnish. *J. Psycholinguist. Res., 31*(2), 83–106.

Vannest, J., & Boland, J. (1999). Lexical morphology and lexical access. *Brain Lang., 68*(1–2), 324–332.

Witt, K., Nühsman, A., & Deuschl, G. (2002). Intact artificial grammar learning in patients with cerebellar degeneration and advanced Parkinson's disease. *Neuropsychologia, 40*(9), 1534–1540.

Wulfeck, B., Bates, E., & Capasso, R. (1991). A crosslinguistic study of grammaticality judgments in Broca's aphasia. *Brain Lang., 41*(2), 311–336.

54 Ventral and Dorsal Contributions to Word Reading

LAURENT COHEN AND STANISLAS DEHAENE

ABSTRACT The core component of expert reading is the fast and accurate perception of single words by the visual system, an ability that results from years of intensive learning. We propose an integrated view of the contributions of the ventral and dorsal streams to this process, associating brain imaging in normal subjects and studies of brain-damaged patients. Together, these two sources of data indicate that fluent reading results from a tight collaboration of both pathways. In the left occipitotemporal cortex, the Visual Word Form system allows for the fast, invariant, and parallel encoding of well-formed letter strings. The occipitoparietal pathway makes an important contribution to reading through attention orienting, word selection, and within-word serial decoding under nonoptimal reading conditions.

The acquisition of reading by children rests on a delicate tuning of the visual system and of the verbal system, and on the elaboration of novel interactions between these two preexisting domains. As a result of this long and effortful process, adult readers are able to scan pages of text in a fast and orderly manner, identifying a flow of words that are each fixated only for a fraction of a second, immediately accessing their sound and meaning, and building up at the same time an integrated interpretation of the text. The core component of this remarkable process is the fast and accurate perception of single words by the visual system. A prerequisite for access to a word's sound and meaning is the identification of its component letters and of their order, an abstract representation that has been called the Visual Word Form (Besner, 1989; Paap, Newsome, & Noel, 1984; Warrington & Shallice, 1980).

In past years, research has concentrated on the contribution of the left ventral visual system to word-identification processes. However, like any complex visual task, reading is most likely achieved through a collaboration of the two components of the cerebral visual system—that is, the ventral occipitotemporal "what" stream and the dorsal occipitoparietal "where" stream (Ungerleider & Mishkin, 1982). In this chapter we propose an integrated view of the contributions of the ventral and dorsal streams to single-word reading. We systematically associate information from brain imaging in normal subjects and contributions from studies of brain-damaged patients with varieties of acquired "peripheral" dyslexias—that is, reading deficits resulting from impaired visual processing, as opposed to language-related "central" dyslexias. Together, these two sources of data indicate that fluent reading results from a tight collaboration of the ventral and dorsal visual pathways, with the occipitotemporal route dominating for expert reading of known words and the occipitoparietal pathway making an essential contribution to reading under dysfluent, unfamiliar, or degraded conditions.

LAURENT COHEN AP-HP, Hôpital de la Salpêtrière, Department of Neurology, Paris; Université Paris VI, Faculté de Médecine Pitié-Salpêtrière, Paris; INSERM UMRS 975, Centre de Recherche de l'ICM, Paris, France

STANISLAS DEHAENE INSERM, Cognitive Neuro-Imaging Unit, Gif sur Yvette; Collège de France, Paris, France

Word processing in the ventral visual pathway

WORD PERCEPTION AS OBJECT PERCEPTION Over the last decades, studies in monkeys and, more recently, functional imaging in humans have shown that object recognition is achieved through neuronal hierarchies located in the ventral occipitotemporal pathway (figure 54.1). Moving from area V1 to inferotemporal (IT) cortex, converging neurons show an increasing invariance to position and scale, an increasing size of the receptive fields, and an increasing complexity of the neurons' optimal stimuli (M. Booth & Rolls, 1998; Riesenhuber & Poggio, 1999; Rolls, 2000; Serre, Oliva, & Poggio, 2007; Ullman, 2007). Connections include bottom-up and top-down projections within the ventral stream (Felleman & Van Essen, 1991), as well as projections to and from more remote frontal and parietal regions subserving attentional control (Kastner & Ungerleider, 2000).

We proposed that the ability to read words stems from this general ability of the ventral stream to identify complex multipart objects. According to the local combination detector, or LCD, model (Dehaene, Cohen, Sigman, & Vinckier, 2005), words are encoded through a posterior-to-anterior hierarchy of neurons tuned to increasingly larger and more complex word fragments, such as visual features, single letters, bigrams, quadrigrams, and possibly whole words.

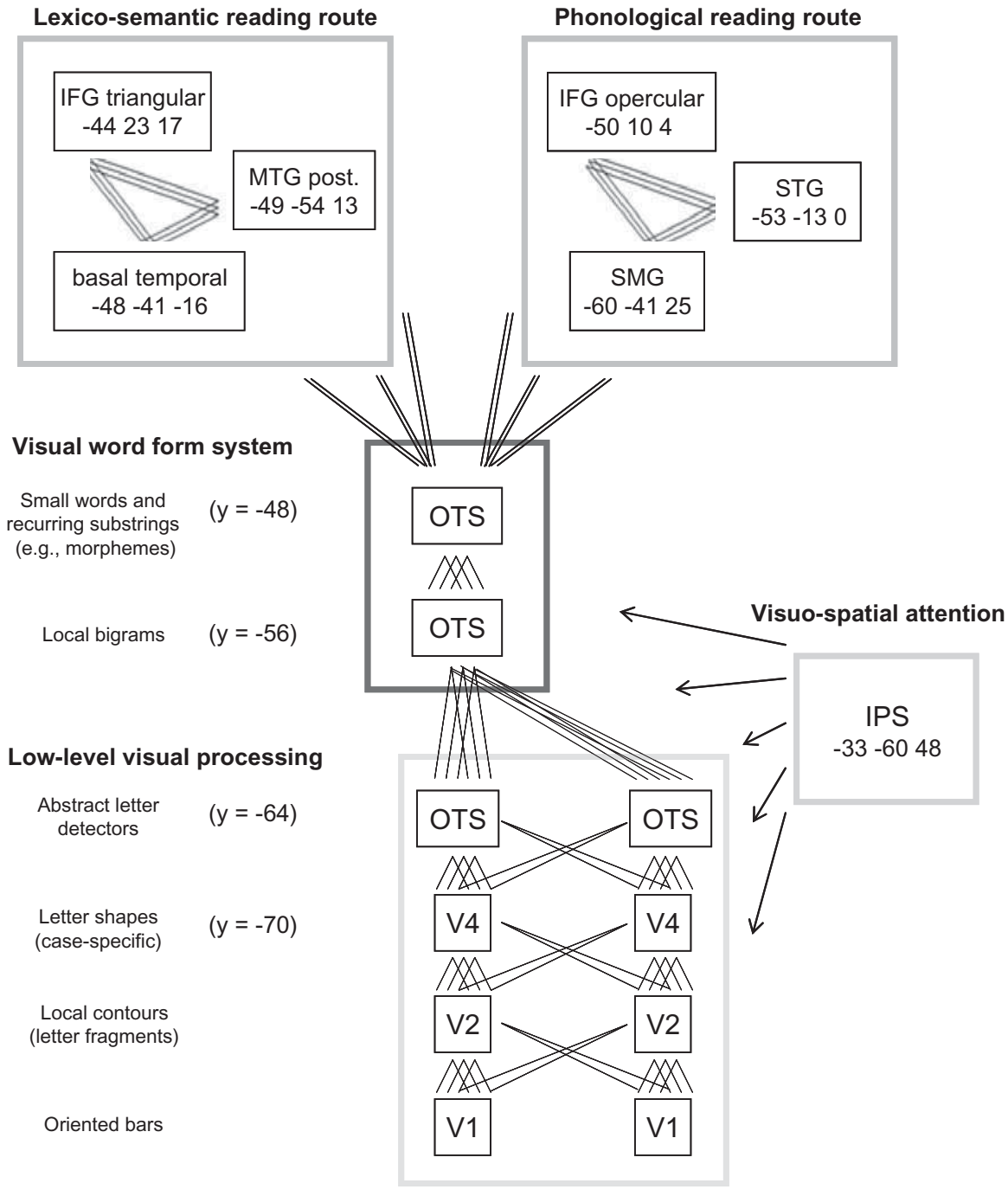


FIGURE 54.1 Synthetic schema of the reading system, merging propositions from Dehaene, Cohen, Sigman, and Vinckier (2005) and Cohen and colleagues (2003). Low-level processing is achieved in each hemisphere for the contralateral half of the visual field (yellow). Information converges on the left-hemispheric Visual Word Form system, where an invariant representation of letter strings is computed (red). The dorsal visual stream exerts a top-down attentional control on the hierarchy of ventral areas (blue). The ventral visual system then feeds the lexicosemantic and phonological reading routes (green). The proposed normalized coordinates for the lexicosemantic and phonological reading routes are from a meta-analysis of 35 PET and fMRI studies (Jobard, Crivello, & Tzourio-Mazoyer, 2003), and the coordinates of the visuospatial attention system are from Gitelman et al. (1999). IFG: inferior frontal gyrus; MTG: middle temporal gyrus; SMG: supramarginal gyrus: OTS: occipitotemporal sulcus; IPS: intraparitetal sulcus. (See color plate 68.)

This system reaches its optimal level of expertise only after years of practice. Through perceptual learning mechanisms, neurons within the ventral pathway become progressively attuned to the regularities of the writing system at all hierarchical levels. This hierarchy must also take into account the need to interact with downstream codes for phonological, morphological, and lexical knowledge of words (Goswami & Ziegler, 2006). Eventually, the adult pattern of performance—that is, fast and invariant word recognition with little influence of the number of letters—is thought to reflect the parallel encoding of letter strings through a fast bottom-up hierarchy of converging detectors.

Early Visual Processing of Printed Words

Retinotopic processing Letters are first processed in the hemisphere contralateral to their location in the visual field, probably in increasingly invariant format, through areas V1 to V4. Those areas, located approximately between Talairach coordinates (TC) $y = -90$ and $y = -70$, are modulated by physical parameters such as word length (Whiting et al., 2003) and visual contrast (Mechelli, Humphreys, Mayall, Olson, & Price, 2000), stimulus degradation (Helenius, Tarkiainen, Cornelissen, Hansen, & Salmelin, 1999; Jernigan et al., 1998), and stimulus rate and duration (Price & Friston, 1997; Price, Moore, & Frackowiak, 1996). Accordingly, the P150 wave evoked by word reading is only sensitive to the physical repetition of stimuli in a masked priming paradigm (Petit, Midgley, Holcomb, & Grainger, 2006).

Perceptual asymmetry It has long been recognized that words are read more easily when they are displayed in the right visual field (RVF) than in the left visual field (LVF) (for reviews see Ducrot & Grainger, 2007; Ellis, 2004). By continuously varying fixation point inside and outside words, Brysbaert, Vitu, and Schroyens (1996) showed that the RVF advantage is closely related to another behavioral asymmetry, namely, that in the optimal reading position, gaze position falls left of word center (Nazir, 2000; O'Regan, Levy-Schoen, Pynte, & Brugaillere, 1984), so that most of the word falls in the RVF. Thus the visual reading span of about 10 letters (Rayner & Bertera, 1979) is not distributed equally across both hemifields, as letter-identification performance decreases more slowly with eccentricity in the RVF than in the LVF (Nazir, Jacobs, & O'Regan, 1998). In addition to higher accuracy and shorter latencies, the RVF advantage is characterized by parallel letter identification, as indexed by constant reading latencies irrespective of word length. The absence of a word-length effect is restricted to words displayed in the optimal viewing position, or fully within the sector of the RVF closest to the fovea. Outside of those conditions, a length effect emerges. Accordingly, when words extend across central fixation, only their left part induces a length effect (Lavidor & Ellis, 2002; Lavidor, Ellis, Shillcock, & Bland, 2001).

The RVF advantage is a complex phenomenon, for which several compatible mechanisms have been put forward: degradation of information resulting from right-to-left interhemispheric transfer of LVF letters; better perceptual learning in the most stimulated sector of the visual field (Nazir, 2000; Nazir, Ben-Boutayab, Decoppet, Deutsch, & Frost, 2004); and rightward attentional bias. As to the ultimate causes of such perceptual or attentional asymmetries, they may involve left-hemispheric lateralization of language (M. Kinsbourne, 1972), left-to-right reading habits (Deutsch & Rayner, 1999; Lavidor & Whitney, 2005; Mishkin & Forgays, 1952), and the fact that the beginning of words is more informative than their end and should therefore be kept close to fixation, as acuity drops steeply away from the fovea (e.g., O'Regan et al., 1984).

Nazir and colleagues (Nazir, 2000; Nazir et al., 2004) emphasized the role of perceptual learning in the genesis of the RVF advantage, as a result of the most frequent perception of words in this sector of the visual field. Along those lines, it is plausible that expert word perception, like other instances of overpracticed perceptual abilities, is restricted to the trained region of the visual field and results from increased activation in retinotopic cortex, with increasing reliance on its more posterior sectors (Sigman et al., 2005). Congruent with this view, Cohen and colleagues (2002) found a left extrastriate region (TC −24 −78 −12) only responsive to RVF stimuli, which showed stronger activation by alphabetic strings than by checkerboards, while no such difference was observed in corresponding right extrastriate areas. Moreover, transcranial magnetic stimulation (TMS) inhibition of the left (but not of the right) occipital cortex induces a length effect for words displayed in the RVF (Skarratt & Lavidor, 2006). This effect occurs when TMS is applied 80 ms after word presentation, supporting the localization of the interference to the posterior visual cortex.

Moreover, priming tasks with split-field stimuli suggest that alphabetic strings are encoded in a format less dependent on physical shape and case when they are viewed in the RVF than in the LVF (Burgund & Marsolek, 1997; Marsolek, Kosslyn, & Squire, 1992; Marsolek, Schacter, & Nicholas, 1996), possibly reflecting general processing asymmetries in the visual system (Burgund & Marsolek, 2000; Marsolek, 1995; Sawamura, Georgieva, Vogels, Vanduffel, & Orban, 2005). Accordingly, using a masked priming paradigm, Dehaene and colleagues (2001) have evidenced case-specific physical repetition priming in the right extrastriate cortex (though similar regions were also present in left extrastriate at a lower threshold) (for similar effects with object perception see Koutstaal et al., 2001).

Overall, such data support the idea that the posterior sector of the left ventral pathway develops superior

perceptual abilities for contralateral strings of letters (as indexed by measures of accuracy, speed, parallelism, and invariance), explaining at least the perceptual component of the RVF advantage.

Pathology: Reading with hemianopia or with apperceptive agnosia The asymmetric role of posterior visual cortex in reading is supported by the pattern of reading impairments resulting from left versus right hemianopia. Reading is highly dependent on the integrity of the central visual field. As unilateral lesions affecting the retrochiasmatic visual tract up to primary visual cortex result in scotomas sparing at least half of the fovea, the ensuing reading impairments are relatively mild. Only right hemianopia without sparing of foveal vision induces noticeable reading difficulty (Zihl, 1995). First, the visual span of such patients is reduced, and they may require several fixations in order to perceive long words. Second, patients lose the reading advantage specific to the normal RVF. Accordingly, they show an influence of word length on reading latencies, as normal subjects do with words displayed in their LVF (Cohen et al., 2003). Third, perception in the right parafoveal field, in an area spanning about 15 letters (Rayner & McConkie, 1976), is important for preparing the accurate landing of the gaze on subsequent words (Sereno & Rayner, 2003). Therefore hemianopic patients make abnormally short and numerous saccades when reading word sequences (Leff et al., 2000; Zihl, 1995).

Finally, patients with so-called apperceptive agnosia (Humphreys & Riddoch, 1993; Lissauer, 1890) following (generally bilateral) lesions of intermediate visual areas such as V2 and V4 are impaired at word reading just as they are at identifying other types of shapes and objects (Heider, 2000; Michel, Henaff, & Bruckert, 1991; Rizzo, Nawrot, Blake, & Damasio, 1992).

Invariant Representation of Letters and the Visual Word Form Area After percolating through retinotopic cortex, visual word information converges on the sector of ventral cortex anterior to V4, ranging approximately from TC $y = 60$ to $y = -40$, a region with larger receptive fields and greater capacity of invariance (figure 54.2). This region receives afferences from both visual hemifields (Tootell, Mendola, Hadjikhani, Liu, & Dale, 1998) and shows repetition suppression by object images across changes in size, position, and orientation (Grill-Spector et al., 1999), and across a change of exemplar within a category (Koutstaal et al., 2001). Accordingly, we proposed that, during reading, part of this region (which we labeled as the Visual Word Form Area, or VWFA) is responsible for the computation of an invariant representation of letter identities (Cohen et al., 2000). Both this proposed labeling and the functional properties of this region have given rise to enduring controversies (Price & Devlin, 2003; Wright et al., 2007), which we tried to clarify by applying to the VWFA the distinctive notions of reproducible localization, partial regional selectivity, and functional specialization (for review and discussion see Cohen & Dehaene, 2004).

Specialization within the ventral stream

1. *Reproducible localization*. Reading-related activations are reproducibly located within the occipitotemporal sulcus lateral to the left fusiform gyrus (VWFA), with only a few millimeters of intersubject variability (Cohen et al., 2002; Jobard, Crivello, & Tzourio-Mazoyer, 2003). The VWFA is activated by visual words irrespective of their position in the visual field (Cohen et al., 2000). An associated electrical or magnetic signature is detected about 170–200 ms after stimulation (e.g., Cohen et al., 2000; Marinkovic et al., 2003; Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999).

The remarkable topographical reproducibility of the VWFA may result from its optimal positioning within gradients biasing the a priori organization of the visual cortex, such as a posterior-to-anterior increase in perceptual invariance (Grill-Spector et al., 1998; Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001) and a mesial-to-lateral increase in preference for foveal versus peripheral stimuli (Hasson, Levy, Behrmann, Hendler, & Malach, 2002). A further reason for the localization of the VWFA, particularly for its usual left lateralization, may be the availability of more direct connections to other language-related sites involved in phonological or lexical processing (Cai, Lavidor, Brysbaert, Paulignan, & Nazir, 2008; Cohen, Jobert, Le Bihan, & Dehaene, 2004; Epelbaum et al., 2008; Mahon & Caramazza, 2009).

2. *Partial regional selectivity*. The VWFA is activated by alphabetic strings relative to fixation but often also relative to complex nonalphabetic stimuli such as faces or geometrical patterns (e.g., Cohen et al., 2002; Puce, Allison, Asgari, Gore, & McCarthy, 1996). However, the difference in activation between words relative to visual objects is variable across studies, and may even be inverted, depending on a number of experimental parameters (e.g., Wright et al., 2007). This lack of absolute regional selectivity may be taken as a sensible argument against the use of the VWFA label, as this region may well be involved in processing nonalphabetic visual objects. However, selectivity may be detectable only at a higher spatial resolution. Thus intracranial recordings occasionally showed P150 or N200 waves elicited exclusively by letter strings, as compared to a variety of control stimuli such as phase-scrambled strings, flowers, faces, or geometrical shapes (Allison, McCarthy, Nobre, Puce, & Belger, 1994; Allison, Puce, Spencer, & McCarthy, 1999). Moreover, some left inferotemporal lesions (see the subsection "Pathology: Pure alexia") yield massive alexia

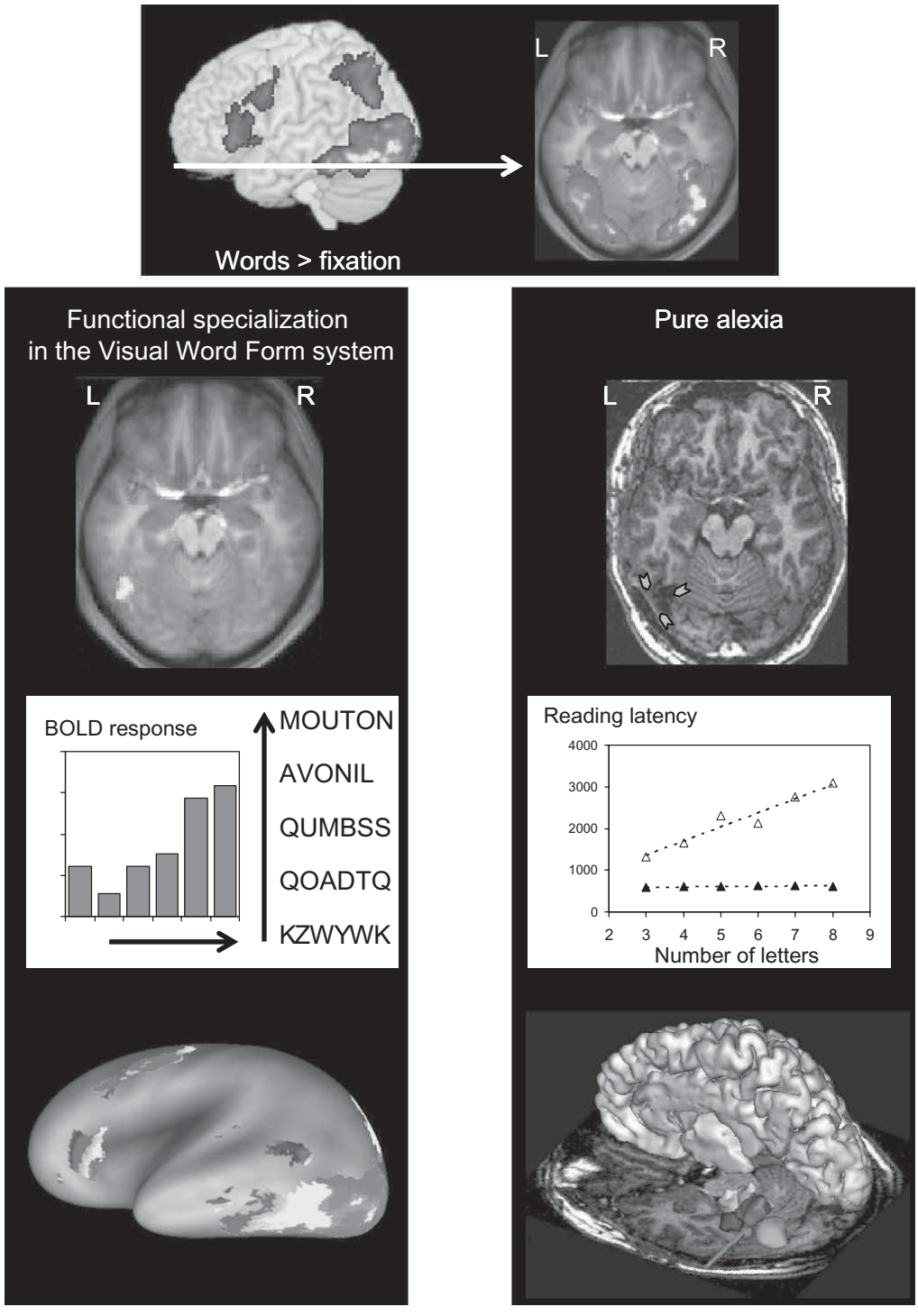


FIGURE 54.2 Word processing in the ventral pathway. *Top panel*: Activations induced by printed words relative to a fixation baseline in the left hemisphere (left) and in the bilateral ventral visual pathway (right). *Left panel*: The VWF system shows a linear increase of activation (top) by letter strings forming closer statistical approximations to orthographically legal strings (middle). This functional specialization increases progressively in more anterior regions within the VWF system (bottom). (Left panel adapted from Vinckier et al., 2007.) *Right panel*: Surgical lesion in the left ventral cortex responsible for pure alexia (top). Whereas before surgery word reading was fast and constant irrespective of word length, after surgery the patient showed slow letter-by-letter reading (middle). In the same patient, the 3D image shows the relative position of the VWFA (blue), of other category-dependent fMRI activation clusters before surgery, of the brain lesion (green), and of intracerebral electrodes (magenta). (Right panel adapted from Gaillard et al., 2006.) (See color plate 69.)

affecting even single letters, contrasting with the spared recognition of complex multipart objects, faces, or digit strings, demonstrating that the VWFA, even if activated by a wide range of stimuli, may evolve to be *necessary* only to word recognition.

3. *Functional specialization.* The issue of selectivity is independent of the hypothesis of a functional specialization of the VWFA (figure 54.2). On top of their preexisting object coding properties, neurons in the VWFA develop elaborate functional specialization as they get attuned to arbitrary features of the subject's script. As the clearest instance of functional specialization, activation of the VWFA is stronger when the script is familiar than when it is unfamiliar (e.g., Hebrew versus alphabetic strings; Baker et al., 2007) or created de novo (Price, Wise, & Frackowiak, 1996). Moreover, using masked repetition priming, it was shown that the VWFA represents words in a format invariant for the upper- versus lowercase distinction (e.g., radio versus RADIO), another arbitrary culture-dependent feature of writing systems (Dehaene et al., 2004, 2001). Finally, within the subjects' familiar script, the VWFA is activated more strongly by letter strings forming closer statistical approximations to orthographically legal strings (including real words), showing that the VWFA incorporates constraints on letter combinations, which are specific to the familiar language (Binder, Medler, Westbury, Liebenthal, & Buchanan, 2006; Cohen et al., 2002; Vinckier et al., 2007).

4. *Internal structure of the Visual Word Form system.* According to the LCD model, the anteroposterior extension of the VWFA (about 20 mm) should reflect its heterogeneous and hierarchically organized structure. Dehaene and colleagues (2004), using a subliminal priming design, showed that the type of prime-target similarity that causes fMRI priming varies according to the anterior-posterior location in left occipitotemporal cortex, with an increasing invariance for position and case change, and probably greater reliance on larger-size units such as bigrams or quadrigrams. More recently, Vinckier and colleagues (2007) tested whether a hierarchy of detectors of increasingly larger word fragments is present in the left occipitotemporal cortex. The frequency of letters, bigrams, and quadrigrams was manipulated, yielding a range of stimuli with an increasing structural similarity to real words. The more anterior an area was within the Visual Word Form region, the more sensitive it was to the frequency of complex components, revealing a gradient-like spatial organization *within* the VWFA (see Grainger and Holcomb, in press, for a review of ERP data relevant to the fragmentation of orthographic processing).

Pathology: Pure alexia Impairments affecting the Visual Word Form system correspond to the syndrome of pure alexia, as described in the 19th century (Binder & Mohr, 1992; Damasio & Damasio, 1983; Dejerine, 1892; see figure 54.2). Pure alexia is an acquired and selective reading deficit occurring in previously literate patients. Patients typically have entirely preserved production and comprehension of oral language, and they can write normally either spontaneously or to dictation. However, they show various degrees of impairment of word reading. The critical cortical lesions generating pure alexia overlap with the VWFA as defined with functional imaging (Cohen et al., 2003; Gaillard et al., 2006). Pure alexia may also follow deafferentation of an intact VWFA following left-hemispheric white matter lesions (Cohen, Henry, et al., 2004; Epelbaum et al., 2008). Posterior callosal lesions cause a selective deafferentation of the VWFA from the right occipital cortex, yielding alexia restricted to the LVF (Cohen et al., 2000, 2003; Molko et al., 2002; Suzuki et al., 1998).

In the most severe cases, known as global alexia, patients cannot identify single letters, let alone whole words (Dalmas & Dansilio, 2000; Dejerine, 1892). Such patients may or may not have access to abstract letter identities, as tested for instance in a cross-case letter-matching task (Miozzo & Caramazza, 1998; Mycroft, Hanley, & Kay, 2002). More often, patients show relatively preserved letter identification abilities and develop letter-by-letter reading strategies, as if only the most finely tuned mechanisms of word perception were affected, those allowing for rapid and parallel identification of letter strings. As an indication of this effortful reading strategy, patients show a large increase in the number and the duration of fixations per word relative to normals and even to patients with hemianopic dyslexia (Behrmann, Shomstein, Black, & Barton, 2001). There is some evidence that in letter-by-letter readers, residual letter identification can be subtended by right-hemispheric regions symmetrical to the VWFA or by spared patches of left-hemispheric ventral cortex (Bartolomeo, Bachoud-Levi, Degos, & Boller, 1998; Cohen, Henry, et al., 2004; Gaillard et al., 2006).

Finally, some patients show better-than-chance performance in purely implicit reading tasks such as lexical or semantic decision, contrasting with the apparent inability to identify printed words (Coslett & Saffran, 1989; Coslett, Saffran, Greenbaum, & Schwartz, 1993). Implicit reading has been most clearly evidenced with Arabic numerals, which can be compared accurately even when explicit reading is grossly impaired (Cohen & Dehaene, 1995, 2000), probably revealing effective right-hemispheric identification processes.

Contribution of the dorsal pathway

The operation of the ventral stream during word reading is modulated by attentional influences, originating from parietal regions, that may impinge on all processing levels from striate cortex (Chawla, Rees, & Friston, 1999; Somers, Dale,

Seiffert, & Tootell, 1999) to ventral occipitotemporal areas (Kastner, De Weerd, Desimone, & Ungerleider, 1998; see figures 54.1 and 54.3). In order to make sense of the variety of reading impairments that may follow parietal lesions, we will distinguish somewhat artificially three contributions of attentional control to single-word reading: orienting to the region of space where the target word is displayed, filtering out irrelevant words present in the vicinity of the target, and serially attending to letters or word fragments whenever letters cannot be effectively processed in parallel over the whole string.

Orientation of Attention Spatial attention modulates the efficiency of the visual processing of alphabetic stimuli. Thus words are better recognized when they appear in a region of the visual field to which attention has been directed by a previous cue (McCann, Folk, & Johnston, 1992), and subliminal letters have a priming effect on subsequent

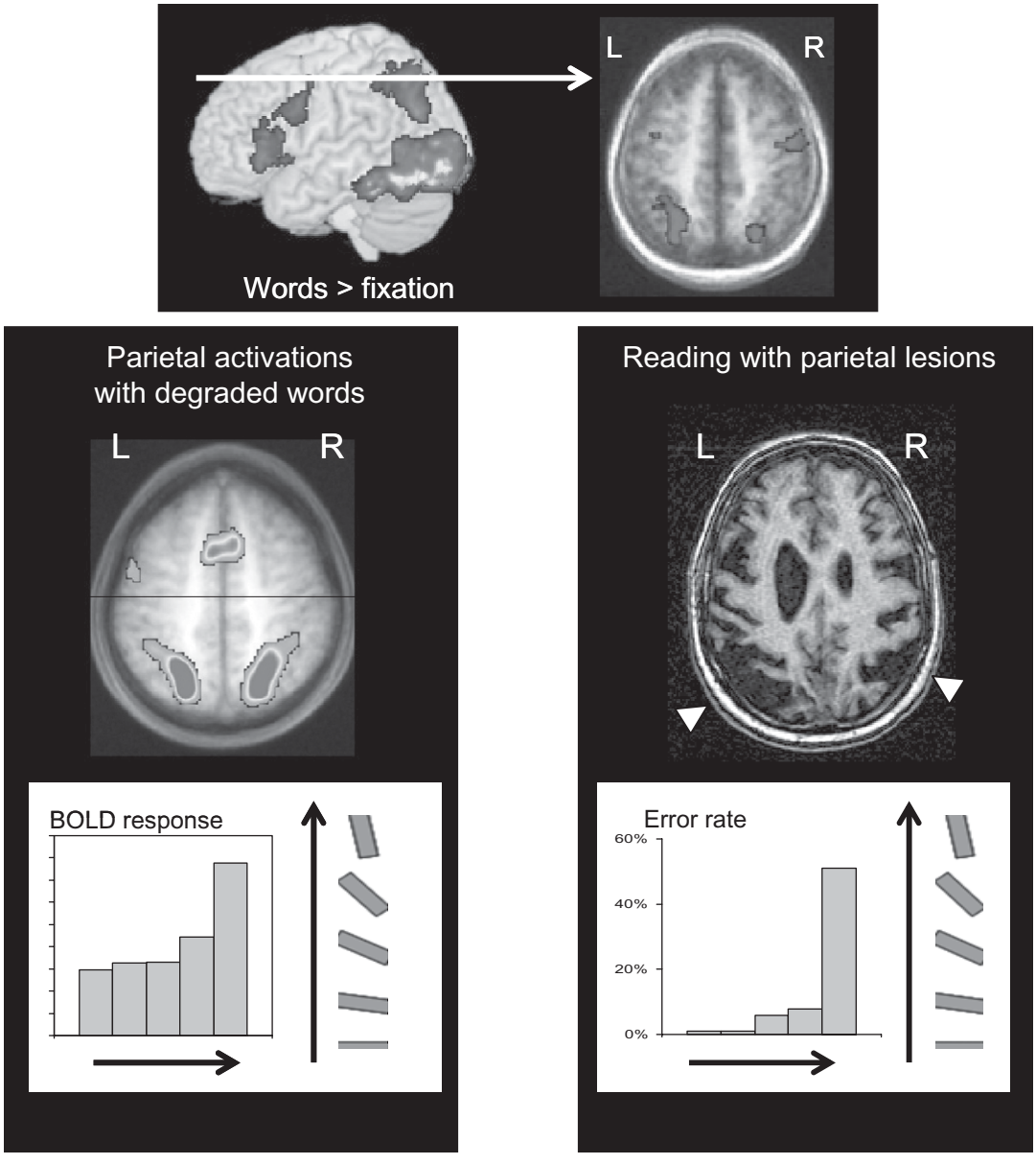


Figure 54.3 Contribution of the dorsal pathway to word reading. *Top panel*: Activations induced by printed words relative to a fixation baseline in the left hemisphere (left) and in the bilateral dorsal visual pathway (right). *Left panel*: The bilateral intraparietal cortex shows a nonlinear increase of activation with word degradation, correlated with reaction times (top). For instance, activations increased steeply for words rotated by more than 45° (bottom). (Left panel adapted from Cohen, Dehaene, Vinckier, Jobert, & Montavont, 2008.) *Right panel*: In a patient with bilateral parietal atrophy and spared ventral cortex (top), there was a severe reading impairment above a similar threshold of rotation angle, demonstrating the role of parietal cortex whenever display degradation exceeds the range of invariance in the ventral cortex. (Right panel adapted from Vinckier et al., 2006.) (See color plate 70.)

targets only when they are displayed at an attended location (Marzouki, Grainger, & Theeuwes, 2007). As mentioned before, the RVF advantage may partly result from a rightward bias of attention. Ducrot and Grainger (2007) showed that exogenous spatial cuing has no impact on the (asymmetrical) reading performance for words displayed only slightly off fixation, suggesting that in the central field, the RVF advantage is mostly perceptual. In contrast, cuing was very effective for more peripheral words and tended to reduce the RVF advantage. In a study of lateralized word reading, Cohen and colleagues (2002) found larger activations for RVF than for LVF words in the left precuneus and thalamus, with no activations for the opposite contrast, likely reflecting the attentional component of the RVF advantage.

Pathology: Neglect dyslexia The defining feature of neglect dyslexia is the existence of a left-right spatial gradient in the rate of reading errors far exceeding the normal RVF advantage (for an overview and references see Riddoch, 1990). Following the general pattern of hemispatial neglect, it is much more common to observe left than right neglect dyslexia, although a number of right-sided cases have been reported. Neglect dyslexia is generally associated with signs of neglect outside the domain of reading, although patients with seemingly isolated neglect dyslexia have been reported. Neglect is thought to result from associated impairments of both nonlateralized and lateralized components of attentional/spatial processing (Husain & Rorden, 2003). The latter may depend on saliency maps of the opposite hemispace subtended by each posterior parietal lobe (Medendorp, Goltz, Vilis, & Crawford, 2003; M. Sereno, 2001). Assuming that those lateralized maps contribute to the top-down modulation of the ventral visual stream, one may expect that distinct varieties of neglect dyslexia may arise, depending on the side of the lesion, the affected parietal structure, the ventral regions that are deprived of attentional modulation, and so on. Indeed, there are numerous clinical observations to illustrate this fractionation of neglect dyslexia (Riddoch, 1990).

Neglect errors typically affect the leftmost letters when patients read single words, and the leftmost side of the page when they read connected text. However, those two types of errors can be to some extent doubly dissociated, suggesting that neglect dyslexia is not a homogeneous syndrome (Costello & Warrington, 1987; Kartsounis & Warrington, 1989). This fractionation is best illustrated by the case of patient JR, who suffered from bilateral occipitoparietal lesions (Humphreys, 1998). When presented with words scattered on a page, he omitted the rightmost words, but his reading errors affected the leftmost letters of the words that he picked out. Likewise, he showed left neglect when he was asked to read single words, while he showed right neglect when trying to name the component letters of the same stimuli. This pattern suggests that JR's left lesion yielded right neglect in situations of competition between objects, while his right lesion yielded left neglect in situations of competitions between the parts of an object.

A clarifying framework was proposed by Hillis and Caramazza (1995), who suggested that the varieties of neglect dyslexia may be attributed to spatial attentional biases acting on one or more of progressively more abstract word representations derived from Marr's theory of object perception (Marr, 1982): a peripheral retinocentric feature representation, a stimulus-centered letter-shape level, and a word-centered graphemic representation akin to the Visual Word Form (for a review of supportive data see Haywood & Coltheart, 2000). Thus, in a deficit at the retinocentric level, error rate for a given letter should depend on its position in the visual field relative to central fixation and not on its rank within the target word. In contrast, in a deficit at the stimulus-centered level, error rate should depend on the distance from the center of the word irrespective of the position of the word in the visual field. Naturally, both parameters may be relevant in some if not in the majority of patients. More remote from neglect in its usual sense, neglect at the graphemic level yields errors affecting one end of words irrespective of their spatial position or orientation. Thus patient NG made errors with the last letters (e.g., hound → house) when reading standard words, but also vertical words and mirror-reversed words, as well as when naming orally spelled words and when performing other lexical tasks such as spelling (Caramazza & Hillis, 1990). Note, however, that there are alternative accounts of word-centered neglect dyslexia, in frameworks that refute the existence of object-centered neural representations (Deneve & Pouget, 2003; Mozer, 2002).

Finally, letter strings that are neglected in explicit reading tasks may nevertheless be processed to higher representation levels. This possibility is suggested by preserved performance in lexical decision (Arduino, Burani, & Vallar, 2003), by the fact that erroneous responses often tend to have the same length as the actual targets (K. Kinsbourne & Warrington, 1962), or by higher error rates observed with nonwords than with real words (Sieroff, Pollatsek, & Posner, 1988). The interpretation of such findings is still debated (Riddoch, 1990), but it is plausible that neglected words can be partially processed in the ventral visual pathway in the absence of conscious awareness, as has also been shown in normal subjects (Dehaene et al., 2001; Devlin et al., 2003) and with other types of visual stimuli such as faces or houses in neglect patients (Rees et al., 2000).

Selection of One Single Word For optimal reading, not only should the attention window encompass the target word, but it should also be narrow enough to exclude other neighboring words. In normal subjects it is possible to force a spread of attention over two words, by briefly

presenting two words side by side, and specifying only afterward which of the two should be reported (Davis & Bowers, 2004; Treisman & Souther, 1986). This procedure degrades performance and induces reading errors that are analogous to those observed in the pathological condition known as attentional dyslexia (for qualifications to this analogy see Davis & Bowers, 2004).

Pathology: Attentional dyslexia The hallmark of attentional dyslexia is the contrast between preserved reading of isolated words and high error rates when the target is surrounded by other words (for a review see Davis & Coltheart, 2002). It is generally attributed to an impaired attentional selection of one among several concurrent stimuli (Shallice, 1988). This induces (1) an inaccurate processing of the target (substitutions, additions, or deletions of letters) as a result of the competition by surrounding words and (2) intrusion of distracters into later stages of processing (letter migrations from the flanking words into the response to the target).

Such ideas are in good agreement with imaging data in normals, showing that when multiple objects are presented simultaneously, they exert mutual inhibition, resulting in decreased ventral visual activations (Kastner et al., 1998). Directing attention toward one of the stimuli compensates this reduction of activity. Moreover, the activation induced by distracters in areas T4 and TEO is reduced in proportion to the attention that is paid to the target, and it is inversely correlated with frontoparietal activations (Pinsk, Doniger, & Kastner, 2004). It is thus plausible that in attentional dyslexics, impaired selection abilities, which are unmasked in the presence of flanker words, cause both visual errors due to a weakened representation of the target and letter migrations due to an excessive activation of distracters.

The phenomenon of flanker interference also prevails when patients are asked to read single letters surrounded by other letters. This finding leads to the paradoxical observation that patients may be good at reading isolated words but not at naming their component letters. More generally, interference seems to occur only between items of the same category. In their seminal article Shallice and Warrington (1977) showed that flanking letters but not flanking digits interfered with letter identification. Similarly, there is no mutual interference between letters and whole words (E. K. Warrington, Cipolotti, & McNeil, 1993). One may note that in some patients the interference between letters is the same whether the target and flankers are printed in the same case or not (Shallice & Warrington, 1977; E. K. Warrington et al.), suggesting that the impairment impinges on visual areas that already show high-level invariance, such as the VWFA. Still, the irrelevance of case changes for attentional selection is not absolute. Indeed, letter migrations between words may be reduced by using different typographic cases (Saffran & Coslett, 1996), suggesting that low-level visual features may help to focus the attention on the target word and to discard distracters.

In brief, attentional dyslexia may be due to insufficient attentional focusing on one among several concurrent letters or letter strings represented in the Visual Word Form system. Note that the few cases of attentional dyslexia with sufficient lesion data consistently point to a left parietal involvement (Friedmann & Gvion, 2001; Mayall & Humphreys, 2002; Shallice & Warrington, 1977; E. K. Warrington et al., 1993). Such asymmetry may relate to a left-hemispheric bias for object-oriented attention (Egly, Driver, & Rafal, 1994), or more generally to the left dominance for language.

Attending to Parts of Words and Serial Decoding As an outcome of perceptual learning, in expert readers the ventral visual pathway gets attuned to the perception of normal print: horizontally aligned words presented in the foveal region in a usual font are identified in a fast and parallel manner. There are, however, a number of circumstances in which this optimal encoding is either unavailable or inappropriate to the task at hand, as revealed by slower reading speed and by the emergence of a linear increase of reading latencies with word length. We suggest that this length effect reflects a failure of parallel letter processing in the ventral pathway and indicates the deployment of serial attention to letters or groups of letters (for an alternative account see Whitney, 2001; Whitney & Lavidor, 2004). Serial reading would involve parietal structures driving spatial-attentional processes (Gitelman et al., 1999; Husain & Rorden, 2003; Kanwisher & Wojciulik, 2000; Mesulam, 1999) and a modulation by this top-down attention of ventral occipitotemporal structures coding for word fragments (Chawla et al., 1999; Kastner et al., 1998; Somers et al., 1999).

Departure from parallel reading as indexed by the emergence of a length effect occurs in many conditions: (1) in children whose reading expertise is still incompletely developed, with an effect of word length persisting until about the age of 10 (Aghababian & Nazir, 2000); (2) in pure alexic patients who develop letter-by-letter reading following left ventral lesions, a strategy that is associated with parietal activations (Gaillard et al., 2006); (3) in normal subjects attempting to read words degraded by means of contrast reduction (Legge, Ahn, Klitz, & Luebker, 1997), of mIxEd case printing (Lavidor, 2002; Mayall, Humphreys, Mechelli, Olson, & Price, 2001), of vertical display (Bub & Lewine, 1988), and of lateral display in the LVF (Lavidor & Ellis, 2002); and (4) in normal subjects reading aloud pseudowords, which probably requires the serial left-to-right conversion of graphemes into phonemes (Weekes, 1997). Interestingly, patients with semantic dementia who suffer from a progressive dissolution of lexical knowledge show a length effect even when reading real words (Cumming,

Patterson, Verfaellie, & Graham, 2006). This abnormal length effect is due to reduced top-down lexical support for word identification, compelling patients to process real words as pseudowords.

We recently studied the mechanisms involved in reading degraded words (Cohen, Dehaene, Vinckier, Jobert, & Montavont, 2008; see figure 54.3). We presented adult readers with words that were progressively degraded in three different ways (word rotation, letter spacing, and displacement to the visual periphery). Behaviorally, we identified degradation thresholds above which reading difficulty increased nonlinearly, with the concomitant emergence of a length effect. Functional MRI activations were correlated with reading difficulty in bilateral occipitotemporal and parietal regions, reflecting the strategies required to identify degraded words. A core region of the intraparietal cortex was engaged in all modes of degradation. Supporting the current interpretation, the same region is also activated, and its interactions with other parts of the reading network increase, when subjects are required to pay attention to letters within nondegraded words (Bitan et al., 2005; J. Booth et al., 2002). Furthermore, in the ventral pathway, word degradation led to an amplification of activation in the posterior Visual Word Form Area at a level thought to encode single letters. We also found an effect of word length restricted to highly degraded words in bilateral occipitoparietal regions.

Pathology: Spatial dyslexia and Balint's syndrome Balint's syndrome, a consequence of bilateral dorsal parietal lesions, includes simultanagnosia, which prevents the binding of objects with a stable localization in space and the computation of their relative positions, and ocular apraxia, which precludes an accurate control of saccades toward peripheral targets (Rizzo & Vecera, 2002). The most salient impact of this disorder on reading is an inability to read connected text as a result of chaotic scanning of the display. The patients' gaze wanders randomly from word to word, and the relative position of words cannot be appreciated. However, patients can read accurately each of the disconnected words on which they land.

While the identification of optimally printed words is not substantially affected, patients may have major difficulties reading words presented in unusual formats, such as vertically arrayed or widely spaced letters. These difficulties disrupt the automatic binding of letters into single visual objects, and therefore require a scanning of component letters, which Balint patients cannot do. Due to impaired scanning, patients may also be unable to report one letter out of a string, even with optimally displayed real words (Baylis, Driver, Baylis, & Rafal, 1994). A similar account explains why Balint patients are impaired at reading pseudowords, for which grapheme-to-phoneme conversion requires the sequential inspection of graphemes. For instance, a patient could read accurately 29 out of 30 briefly presented words, while she identified only 4 out of 30 pseudowords (Coslett & Saffran, 1991).

We recently studied a simultanagnosic patient with bilateral parietal atrophy (Vinckier et al., 2006; see figure 54.3). She was excellent at reading normally printed foveal words, but she was severely impaired at reading words that were mirror reversed, or rotated by angles larger than 50°, or whose letters were separated by at least two blank spaces, or words displayed in her left hemifield. According to the present hypothesis, above those critical thresholds—that is, when stimulus degradation exceeds the perceptual tolerance of the ventral system—reading normally requires the intervention of the parietal lobes to pilot the attention-driven exploration of stimuli (for a congruent observation see Hall, Humphreys, & Cooper, 2001). Parietal lesions did not allow the patient to resort to such strategy. This study was congruent with an imaging study reviewed before (Cohen et al., 2008): overlapping parietal regions were activated in normal subjects and lesioned in the patient, and the same degree of word degradation boosted parietal activations in normals and caused a drop in the patient's performance.

Because of her parietal lesions, this patient also presented with orientation agnosia (e.g., Priftis, Rusconi, Umilta, & Zorzi, 2003). She was thus unable to discriminate normally oriented words or pictures of objects from the same rotated stimuli. However, while she was unable to discriminate pictures of objects from their mirror-reversed images, she could do so easily with reversible pseudowords. For instance, "boup" and "quod" appeared to her as distinct items, although they are mirror images of each other. The ventral pathway builds up a mirror-invariant representation of common objects (Logothetis & Pauls, 1995; Rollenhagen & Olson, 2000), which requires the intervention of explicit orientation analysis dependent on parietal cortex in order to discriminate mirror images. In contrast, the default invariance for mirror symmetry is "unlearned" by the ventral pathway in the particular case of reading, since reading requires the accurate discrimination of mirror-symmetric shapes (e.g., "p" versus "q").

Interfacing with the verbal system

As the result of a collaboration between ventral and dorsal routes, detailed visual information about letter strings is ultimately conveyed to downstream language areas. In this section, we briefly point out some open issues pertaining to the relationships of the visual system with the language-related components of word processing, including phonology and the lexicon.

Multiple Outlets from the Ventral Stream Assuming that word fragments of various sizes are identified in the ventral stream, one may expect that rich direct and indirect projections should exist toward areas involved in lexical, semantic, motor, or phonological processes. However, the pathways leading from the VWFA to all components of the reading network are not precisely defined. The macaque equivalent of the VWFA putatively falls within the IT complex, which projects to the inferior parietal lobule and the anterior temporal lobe, in addition to occipital and interhemispheric connections (Schmahmann & Pandya, 2006). Moreover, there may be a specifically human development of projections from the inferior temporal cortex to language-related superior temporal, parietal, and frontal regions, through the arcuate fasciculus (Catani, Jones, & ffytche, 2005; Epelbaum et al., 2008) and the inferior fronto-occipital fasciculus (Catani, Howard, Pajevic, & Jones, 2002), respectively.

Following the observation of alexia with agraphia, Dejerine (1892) suggested that the next step following visual word processing should be the angular gyrus, which he postulated to be the "visual center of letters." Indeed, the angular gyrus is among the regions that are modulated during reading tasks, even if it often remains below the baseline level of activation (Binder et al., 2003; Binder, Medler, Desai, Conant, & Liebenthal, 2005), and there is functional connectivity between the angular gyrus and the left fusiform gyrus at coordinates matching the VWFA (Horwitz, Rumsey, & Donohue, 1998). There is also correlated activity in the VWFA and in left inferior frontal areas (Bokde, Tagamets, Friedman, & Horwitz, 2001). A further potential output pathway is to temporal regions anterior to the VWFA. These regions, which have been difficult to image with functional MRI because of magnetic susceptibility artifacts, are probably involved in supramodal semantic processing (for a review see Giraud & Price, 2001; Kreiman, Koch, & Fried, 2000; Lambon Ralph, McClelland, Patterson, Galton, & Hodges, 2001).

Finally, it is possible that different segments of the Visual Word Form system feed distinct language-related processes by projecting to distinct areas. Thus Mechelli and colleagues (2005) found that during reading the posterior fusiform cortex, which codes for single letters according to the LCD model, was coupled with the superior premotor cortex, possibly in relation to letter-to-articulation transcoding, while the anterior fusiform cortex, presumably coding for large word fragments, was coupled with Broca's pars triangularis, possibly in relation to lexicosemantic access. Accordingly, the former coupling increased during pseudoword reading, whereas the latter increased during exception word reading. In a similar vein, Grainger proposed on the basis of behavioral data that two types of orthographic code are computed: a coarse code used to rapidly access semantic information and a finer-grained code used to access phonology from orthography (Grainger & Holcomb, in press).

Phonological Impact on Visual Representations One potential shortcoming of the LCD model is that it focuses primarily on the acquisition of visual expertise in reading—that is, how the ventral visual system eventually incorporates orthographic regularities (see figure 54.1). However, it is likely that word phonology also influences orthographic representations in the visual system. Early letter-to-sound mapping is thought to be crucial for reading acquisition, which may constrain the eventual structure of the orthographic code in adults (Goswami & Ziegler, 2006; Ziegler & Goswami, 2005).

The impact of phonology on visual processing emerges from the comparison between scripts that differ in terms of orthographic transparency—that is, the regularity of grapheme-phoneme conversion rules. According to the LCD model, transparency should be reflected in the size of the units encoded by occipitotemporal neurons. In "transparent" writing systems such as Italian or the Japanese kana script, the letter and bigram levels should suffice for grapheme-phoneme conversion. In an "opaque" script, however, such as English or kanji, a larger-size visual unit, more anterior along the visual hiearchy, should be used. Compatible with this idea, stronger and more anterior activation is observed in the left occipitotemporal region in English than in Italian readers (Paulesu et al., 2000), and, at a slightly more mesial location, during kanji than during kana reading in Japanese readers (Ha Duy Thuy et al., 2004; Nakamura, Dehaene, Jobert, Le Bihan, & Kouider, 2005).

However, evidence of an influence of phonology on visual processing within a given writing system is less clear. There are numerous behavioral demonstrations of an impact of phonology on the processing of printed words, as well as of cross-modal word activations in parietal and superior or lateral temporal regions (e.g., J. Booth et al., 2002; Cohen, Jobert, Le Bihan, & Dehaene, 2005; van Atteveldt, Formisano, Goebel, & Blomert, 2004). Still there is little evidence that some of those effects reflect the operation of the visual system per se, rather than of later speech-related processes. For instance, Grainger, Kiyonaga, & Holcomb (2006) showed that by 225 ms after the presentation of a target word preceded by a masked prime, ERPs distinguished homophone pseudoword primes, as compared to nonhomophone controls (e.g., bakon-BACON versus bafon-BACON). Although this time window is roughly compatible with processing in the Visual Word Form system, the anterior topography of this effect does not support an occipito-temporal source. The contribution of phonological structure to word encoding in the visual system is thus largely open to empirical research.

Conclusion

The present review emphasizes that fluent reading results from an intimate collaboration of multiple areas forming a distributed network. Although the VWFA clearly plays an essential role in expert reading, the recent literature has tended to forget that the dorsal spatial-attentional system also makes a major contribution through attention orienting, word selection, and within-word serial decoding. Adult readers probably rely on serial attentive reading under relatively rare conditions; but we speculate that young readers, in whom the word length effect is particularly large, rely heavily on the dorsal route early during the laying down of the grapheme-phoneme decoding stage. Although phonological sources of developmental reading impairments have received vast attention, our analysis suggests that occipito-parietal impairments are also very likely to have an impact on developmental dyslexia, as indeed suggested by recent research (Bosse, Tainturier, & Valdois, 2007; Lassus-Sangosse, N'Guyen-Morel, & Valdois, 2008; Valdois, Bosse, & Tainturier, 2004). In the future, developmental neuroimaging paradigms should be developed to directly image the ventral and dorsal routes as children learn to read.

REFERENCES

Aghababian, V., & Nazir, T. A. (2000). Developing normal reading skills: Aspects of the visual processes underlying word recognition. *J. Exp. Child Psychol., 76*(2), 123–150.

Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb. Cortex, 4*(5), 544–554.

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb. Cortex, 9*(5), 415–430.

Arduino, L. S., Burani, C., & Vallar, G. (2003). Reading aloud and lexical decision in neglect dyslexia patients: A dissociation. *Neuropsychologia, 41*(8), 877–885.

Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc. Natl. Acad. Sci. USA, 104*(21), 9087–9092.

Bartolomeo, P., Bachoud-Levi, A. C., Degos, J. D., & Boller, F. (1998). Disruption of residual reading capacity in a pure alexic patient after a mirror-image right-hemispheric lesion. *Neurology, 50*(1), 286–288.

Baylis, G. C., Driver, J., Baylis, L. L., & Rafal, R. D. (1994). Reading of letters and words in a patient with Balint's syndrome. *Neuropsychologia, 32*(10), 1273–1286.

Behrmann, M., Shomstein, S. S., Black, S. E., & Barton, J. J. (2001). The eye movements of pure alexic patients during reading and nonreading tasks. *Neuropsychologia, 39*(9), 983–1002.

Besner, D. (1989). On the role of outline shape and word-specific visual pattern in the identification of function words—none. *Q. J. Exp. Psychol. [A], 41*, 91–105.

Binder, J. R., McKiernan, K. A., Parsons, M. E., Westbury, C. F., Possing, E. T., Kaufman, J. N., et al. (2003). Neural correlates of lexical access during visual word recognition. *J. Cogn. Neurosci., 15*(3), 372–393.

Binder, J. R., Medler, D. A., Desai, R., Conant, L. L., & Liebenthal, E. (2005). Some neurophysiological constraints on models of word naming. *NeuroImage, 27*(3), 677–693.

Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *NeuroImage, 33*(2), 739–748.

Binder, J. R., & Mohr, J. P. (1992). The topography of callosal reading pathways. A case-control analysis. *Brain, 115*, 1807–1826.

Bitan, T., Booth, J. R., Choy, J., Burman, D. D., Gitelman, D. R., & Mesulam, M. M. (2005). Shifts of effective connectivity within a language network during rhyming and spelling. *J. Neurosci., 25*(22), 5397–5403.

Bokde, A. L., Tagamets, M. A., Friedman, R. B., & Horwitz, B. (2001). Functional interactions of the inferior frontal cortex during the processing of words and word-like stimuli. *Neuron, 30*(2), 609–617.

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2002). Functional anatomy of intra- and cross-modal lexical tasks. *NeuroImage, 16*(1), 7–22.

Booth, M. C., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. Cortex, 8*(6), 510–523.

Bosse, M. L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition, 104*(2), 198–230.

Brysbaert, M., Vitu, F., & Schroyens, W. (1996). The right visual field advantage and the optimal viewing position effect: On the relation between foveal and parafoveal word recognition. *Neuropsychology, 10*, 385–395.

Bub, D. N., & Lewine, J. (1988). Different modes of word recognition in the left and right visual fields. *Brain Lang., 33*(1), 161–188.

Burgund, E. D., & Marsolek, C. J. (1997). Letter-case-specific priming in the right cerebral hemisphere with a form-specific perceptual identification task. *Brain Cogn., 35*(2), 239–258.

Burgund, E. D., & Marsolek, C. J. (2000). Viewpoint-invariant and viewpoint-dependent object recognition in dissociable neural subsystems. *Psychon. Bull. Rev., 7*(3), 480–489.

Cai, Q., Lavidor, M., Brysbaert, M., Paulignan, Y., & Nazir, T. A. (2008). Cerebral lateralization of frontal lobe language processes and lateralization of the posterior visual word processing system. *J. Cogn. Neurosci., 20*(4), 672–681.

Caramazza, A., & Hillis, A. E. (1990). Spatial representation of words in the brain implied by studies of a unilateral neglect patient. *Nature, 346*, 267–269.

Catani, M., Howard, R. J., Pajevic, S., & Jones, D. K. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage, 17*(1), 77–94.

Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol., 57*(1), 8–16.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nat. Neurosci., 2*(7), 671–676.

Cohen, L., & Dehaene, S. (1995). Number processing in pure alexia: The effect of hemispheric asymmetries and task demands. *Neurocase, 1*, 121–137.

Cohen, L., & Dehaene, S. (2000). Calculating without reading: Unsuspected residual abilities in pure alexia. *Cogn. Neuropsychol., 17*, 563–583.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the Visual Word Form Area. *NeuroImage, 22*, 466–476.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., et al. (2000). The Visual Word Form Area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*, 291–307.

Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., & Montavont, A. (2008). Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *NeuroImage, 40*(1), 353–366.

Cohen, L., Henry, C., Dehaene, S., Molko, N., Lehéricy, S., Martinaud, O., et al. (2004). The pathophysiology of letter-by-letter reading. *Neuropsychologia, 42*, 1768–1780.

Cohen, L., Jobert, A., Le Bihan, D., & Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *NeuroImage, 23*(4), 1256–1270.

Cohen, L., Jobert, A., Le Bihan, D., & Dehaene, S. (2005). Distinct unimodal and crossmodal regions for word processing in the left temporal cortex. *NeuroImage, 23*, 1256–1270.

Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain, 125*(Pt. 5), 1054–1069.

Cohen, L., Martinaud, O., Lemer, C., Lehericy, S., Samson, Y., Obadia, M., et al. (2003). Visual word recognition in the left and right hemispheres: Anatomical and functional correlates of peripheral alexias. *Cereb. Cortex, 13*(12), 1313–1333.

Coslett, H. B., & Saffran, E. (1991). Simultanagnosia: To see but not two see. *Brain, 114*, 1523–1545.

Coslett, H. B., & Saffran, E. M. (1989). Evidence for preserved reading in "pure alexia." *Brain, 112*, 327–359.

Coslett, H. B., Saffran, E. M., Greenbaum, S., & Schwartz, H. (1993). Reading in pure alexia: The effect of strategy. *Brain, 116*, 21–37.

Costello, A., & Warrington, E. K. (1987). The dissociation of visuospatial neglect and neglect dyslexia. *J. Neurol. Neurosurg. Psychiatry, 50*, 1110–1116.

Cumming, T. B., Patterson, K., Verfaellie, M., & Graham, K. S. (2006). One bird with two stones: Abnormal word length effect in pure alexia and semantic dementia. *Cogn. Neuropsychol., 23*, 1130–1161.

Dalmas, J. F., & Dansilio, S. (2000). Visuographemic alexia: A new form of a peripheral acquired dyslexia. *Brain Lang., 75*(1), 1–16.

Damasio, A. R., & Damasio, H. (1983). The anatomic basis of pure alexia. *Neurology, 33*, 1573–1583.

Davis, C. J., & Bowers, J. S. (2004). What do letter migration errors reveal about letter position coding in visual word recognition? *J. Exp. Psychol. Hum. Percept. Perform., 30*(5), 923–941.

Davis, C., & Coltheart, M. (2002). Paying attention to reading errors in acquired dyslexia. *Trends Cogn. Sci., 6*(9), 359.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends Cogn. Sci., 9*, 335–341.

Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J. B., Le Bihan, D., et al. (2004). Letter binding and invariant recognition of masked words. *Psychol. Sci., 15*(5), 307–313.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci., 4*(7), 752–758.

Dejerine, J. (1892). Contribution à l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. *Mémoires de la Société de Biologie, 4*, 61–90.

Deneve, S., & Pouget, A. (2003). Basis functions for object-centered representations. *Neuron, 37*, 347–359.

Deutsch, A., & Rayner, K. (1999). Initial fixation location effects in reading Hebrew words. *Lang. Cogn. Process., 14*, 393–421.

Devlin, A. M., Cross, J. H., Harkness, W., Chong, W. K., Harding, B., Vargha-Khadem, F., et al. (2003). Clinical outcomes of hemispherectomy for epilepsy in childhood and adolescence. *Brain, 126*(Pt. 3), 556–566.

Ducrot, S., & Grainger, J. (2007). Deployment of spatial attention to words in central and peripheral vision. *Percept. Psychophys., 69*(4), 578–590.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *J. Exp. Psychol. Gen., 123*(2), 161–177.

Ellis, A. W. (2004). Length, formats, neighbours, hemispheres, and the processing of words presented laterally or at fixation. *Brain Lang., 88*(3), 355–366.

Epelbaum, S., Pinel, P., Gaillard, R., Delmaire, C., Perrin, M., Dupont, S., Dehaene, S., & Cohen, L. (2008). Pure alexia as a disconnection syndrome: New diffusion imaging evidence for an old concept. *Cortex, 44*, 962–974.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex, 1*(1), 1–47.

Friedmann, N., & Gvion, A. (2001). Letter position dyslexia. *Cogn. Neuropsychol., 18*, 673–696.

Gaillard, R., Naccache, L., Pinel, P., Clemenceau, S., Volle, E., Hasboun, D., et al. (2006). Direct intracranial, fMRI and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron, 50*, 191–204.

Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroanatomy places on classical models of auditory word processing. *J. Cogn. Neurosci., 13*, 754–765.

Gitelman, D. R., Nobre, A. C., Parrish, T. B., LaBar, K. S., Kim, Y. H., Meyer, J. R., et al. (1999). A large-scale distributed network for covert spatial attention: Further anatomical delineation based on stringent behavioural and cognitive controls. *Brain, 122*(Pt. 6), 1093–1106.

Goswami, U., & Ziegler, J. C. (2006). A developmental perspective on the neural code for written words. *Trends Cogn. Sci., 10*(4), 142–143.

Grainger, J., & Holcomb, P. J. (in press). Neural constraints on a functional architecture for word recognition. In P. L. Cornelissen, P. C. Hansen, M. L. Kringelbach, & K. Pugh (Eds.), *The neural basis of reading.* Oxford, UK: Oxford University Press.

Grainger, J., Kiyonaga, K., & Holcomb, P. J. (2006). The time course of orthographic and phonological code activation. *Psychol. Sci., 17*(12), 1021–1026.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*(1), 187–203.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum. Brain Mapp., 6*(4), 316–328.

Ha Duy Thuy, D., Matsuo, K., Nakamura, K., Toma, K., Oga, T., Nakai, T., et al. (2004). Implicit and explicit processing of

kanji and kana words and non-words studied with fMRI. *NeuroImage, 23*(3), 878–889.

Hall, D. A., Humphreys, G. W., & Cooper, A. G. C. (2001). Neuropsychological evidence for case-specific reading: Multi-letter units in visual word recognition. *Q. J. Exp. Psychol. [A], 54*, 439–467.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron, 34*(3), 479–490.

Haywood, M., & Coltheart, M. (2000). Neglect dyslexia and the early stages of visual word recognition. *Neurocase, 6*, 33–44.

Heider, B. (2000). Visual form agnosia: Neural mechanisms and anatomical foundations. *Neurocase, 6*, 1–12.

Helenius, P., Tarkiainen, A., Cornelissen, P., Hansen, P. C., & Salmelin, R. (1999). Dissociation of normal feature analysis and deficient processing of letter-strings in dyslexic adults. *Cereb. Cortex, 9*(5), 476–483.

Hillis, A. E., & Caramazza, A. (1995). A framework for interpreting distinct patterns of hemispatial neglect. *Neurocase, 1*, 189–207.

Horwitz, B., Rumsey, J. M., & Donohue, B. C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc. Natl. Acad. Sci. USA, 95*(15), 8939–8944.

Humphreys, G. W. (1998). Neural representation of objects in space: A dual coding account. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 353*(1373), 1341–1351.

Humphreys, G. W., & Riddoch, M. J. (1993). Object agnosias. In C. Kennard (Ed.), *Visual perceptual defects* (pp. 339–359). London: Baillière Tindall.

Husain, M., & Rorden, C. (2003). Non-spatially lateralized mechanisms in hemispatial neglect. *Nat. Rev. Neurosci., 4*, 26–36.

Jernigan, T. L., Ostergaard, A. L., Law, I., Svarer, C., Gerlach, C., & Paulson, O. B. (1998). Brain activation during word identification and word recognition. *NeuroImage, 8*(1), 93–105.

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *NeuroImage, 20*, 693–712.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nat. Rev. Neurosci., 1*(2), 91–100.

Kartsounis, L. D., & Warrington, E. K. (1989). Unilateral neglect overcome by cues implicit in stimulus displays. *J. Neurol. Neurosurg. Psychiatry, 52*, 1253–1259.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science, 282*(5386), 108–111.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci., 23*, 315–341.

Kinsbourne, K., & Warrington, E. K. (1962). A variety of reading dysability associated with right hemisphere lesions. *J. Neurol., 25*, 339–344.

Kinsbourne, M. (1972). Eye and head turning indicates cerebral lateralization. *Science, 176*(34), 539–541.

Koutstaal, W., Wagner, A. D., Rotte, M., Maril, A., Buckner, R. L., & Schacter, D. L. (2001). Perceptual specificity in visual object priming: Functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia, 39*(2), 184–199.

Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat. Neurosci., 3*(9), 946–953.

Lambon Ralph, M. A., McClelland, J. L., Patterson, K., Galton, C. J., & Hodges, J. R. (2001). No right to speak? The relationship between object naming and semantic impairment: Neuropsychological evidence and a computational model. *J. Cogn. Neurosci., 13*(3), 341–356.

Lassus-Sangosse, D., N'Guyen-Morel, M. A., & Valdois, S. (2008). Sequential or simultaneous visual processing deficit in developmental dyslexia? *Vis. Res., 48*, 979–988.

Lavidor, M. (2002). An examination of the lateralized abstractive/form specific model using MiXeD-CaSe primes. *Brain Cogn., 48*(2–3), 413–417.

Lavidor, M., & Ellis, A. W. (2002). Word length and orthographic neighborhood size effects in the left and right cerebral hemispheres. *Brain Lang., 80*(1), 45–62.

Lavidor, M., Ellis, A. W., Shillcock, R., & Bland, T. (2001). Evaluating a split processing model of visual word recognition: Effects of word length. *Brain Res. Cogn. Brain Res., 12*(2), 265–272.

Lavidor, M., & Whitney, C. (2005). Word length effects in Hebrew. *Brain Res. Cogn. Brain Res., 24*(1), 127–132.

Leff, A. P., Scott, S. K., Crewes, H., Hodgson, T. L., Cowey, A., Howard, D., et al. (2000). Impaired reading in patients with right hemianopia. *Ann. Neurol., 47*(2), 171–178.

Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1997). Psychophysics of reading—XVI: The visual span in normal and low vision. *Vis. Res., 37*(14), 1999–2010.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object process-ing stages in the human visual cortex. *Cereb. Cortex, 11*(4), 287–297.

Lissauer, H. (1890). Ein Fall von Seelenblindheit nebst einen Beitrage zur Theorie derselben. *Arch. Psychiatr. Nervenkr., 21*, 222–270.

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cereb. Cortex, 5*(3), 270–288.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annu. Rev. Psychol, 60*, 27–51.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron, 38*(3), 487–497.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman.

Marsolek, C. J. (1995). Abstract visual-form representations in the left cerebral hemisphere. *J. Exp. Psychol. Hum. Percept. Perform., 21*(2), 375–386.

Marsolek, C. J., Kosslyn, S. M., & Squire, L. R. (1992). Form-specific visual priming in the right cerebral hemisphere. *J. Exp. Psychol. Learn. Mem. Cogn., 18*(3), 492–508.

Marsolek, C. J., Schacter, D. L., & Nicholas, C. D. (1996). Form-specific visual priming for new associations in the right cerebral hemisphere. *Mem. Cogn., 24*(5), 539–556.

Marzouki, Y., Grainger, J., & Theeuwes, J. (2007). Exogenous spatial cueing modulates subliminal masked priming. *Acta Psychol. (Amst.), 126*(1), 34–45.

Mayall, K., & Humphreys, G. W. (2002). Presentation and task effects on migration errors in attentional dyslexia. *Neuropsychologia, 40*(8), 1506–1515.

Mayall, K., Humphreys, G. W., Mechelli, A., Olson, A., & Price, C. J. (2001). The effects of case mixing on word recognition: Evidence from a PET study. *J. Cogn. Neurosci., 13*(6), 844–853.

McCann, R. S., Folk, C. L., & Johnston, J. C. (1992). The role of spatial attention in visual word processing. *J. Exp. Psychol. Hum. Percept. Perform., 18*(4), 1015–1029.

Mechelli, A., Crinion, J. T., Long, S., Friston, K. J., Lambon Ralph, M. A., Patterson, K., et al. (2005). Dissociating reading processes on the basis of neuronal interactions. *J. Cogn. Neurosci., 17*(11), 1753–1765.

Mechelli, A., Humphreys, G. W., Mayall, K., Olson, A., & Price, C. J. (2000). Differential effects of word length and visual contrast in the fusiform and lingual gyri during reading. *Proc. R. Soc. Lond. B Biol. Sci., 267*(1455), 1909–1913.

Medendorp, W. P., Goltz, H. C., Vilis, T., & Crawford, J. D. (2003). Gaze-centered updating of visual space in human parietal cortex. *J. Neurosci., 23*, 6209–6214.

Mesulam, M. M. (1999). Spatial attention and neglect: Parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 354*(1387), 1325–1346.

Michel, F., Henaff, M. A., & Bruckert, R. (1991). Unmasking of visual deficits following unilateral prestriate lesions in man. *NeuroReport, 2*(6), 341–344.

Miozzo, M., & Caramazza, A. (1998). Varieties of pure alexia: The case of failure to access graphemic representations. *Cogn. Neuropsychol., 15*, 203–238.

Mishkin, M., & Forgays, D. G. (1952). Word recognition as a function of retinal locus. *J. Exp. Psychol., 43*, 43–48.

Molko, N., Cohen, L., Mangin, J. F., Chochon, F., Lehéricy, S., Le Bihan, D., et al. (2002). Visualizing the neural bases of a disconnection syndrome with diffusion tensor imaging. *J. Cogn. Neurosci., 14*, 629–636.

Mozer, M. C. (2002). Frames of reference in unilateral neglect and visual perception: A computational perspective. *Psychol. Rev., 109*, 156–185.

Mycroft, R., Hanley, J. R., & Kay, J. (2002). Preserved access to abstract letter identities despite abolished letter naming in a case of pure alexia. *J. Neurolinguistics, 15*, 99–108.

Nakamura, K., Dehaene, S., Jobert, A., Le Bihan, D., & Kouider, S. (2005). Subliminal convergence of kanji and kana words: Further evidence for functional parcellation of the posterior temporal cortex in visual word perception. *J. Cogn. Neurosci, 17*, 954–968.

Nazir, T. A. (2000). Traces of print along the visual pathway. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 3–22). Amsterdam: Elsevier.

Nazir, T. A., Ben-Boutayab, N., Decoppet, N., Deutsch, A., & Frost, R. (2004). Reading habits, perceptual learning, and recognition of printed words. *Brain Lang., 88*(3), 294–311.

Nazir, T. A., Jacobs, A. M., & O'Regan, J. K. (1998). Letter legibility and visual word recognition. *Mem. Cogn., 26*, 810–821.

O'Regan, J. K., Levy-Schoen, A., Pynte, J., & Brugaillere, B. (1984). Convenient fixation location within isolated words of different length and structure. *J. Exp. Psychol. Hum. Percept. Perform., 10*(2), 250–257.

Paap, K. R., Newsome, S. L., & Noel, R. W. (1984). Word shape's in poor shape for the race to the lexicon. *J. Exp. Psychol. Hum. Percept. Perform., 10*(3), 413–428.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., et al. (2000). A cultural effect on brain function. *Nat. Neurosci., 3*(1), 91–96.

Petit, J. P., Midgley, K. J., Holcomb, P. J., & Grainger, J. (2006). On the time course of letter perception: A masked priming ERP investigation. *Psychon. Bull. Rev., 13*(4), 674–681.

Pinsk, M. A., Doniger, G. M., & Kastner, S. (2004). Push-pull mechanism of selective attention in human extrastriate cortex. *J. Neurophysiol., 92*(1), 622–629.

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage, 19*(3), 473–481.

Price, C. J., & Friston, K. J. (1997). The temporal dynamics of reading: A PET study. *Proc. R. Soc. Lond. B Biol. Sci., 264*(1389), 1785–1791.

Price, C. J., Moore, C. J., & Frackowiak, R. S. (1996). The effect of varying stimulus rate and duration on brain activity during reading. *NeuroImage, 3*(1), 40–52.

Price, C. J., Wise, R. J. S., & Frackowiak, R. S. J. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cereb. Cortex, 6*, 62–70.

Priftis, K., Rusconi, E., Umilta, C., & Zorzi, M. (2003). Pure agnosia for mirror stimuli after right inferior parietal lesion. *Brain, 126*(Pt. 4), 908–919.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *J. Neurosci., 16*, 5205–5215.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science, 206*(4417), 468–469.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vis. Res., 16*(8), 829–837.

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C., & Driver, J. (2000). Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain, 123*(Pt. 8), 1624–1633.

Riddoch, J. (1990). Neglect and the peripheral dyslexias. *Cogn. Neuropsychol., 7*, 369–389.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci., 2*, 1019–1025.

Rizzo, M., Nawrot, M., Blake, R., & Damasio, A. (1992). A human visual disorder resembling area V4 dysfunction in the monkey. *Neurology, 42*, 1175–1180.

Rizzo, M., & Vecera, S. P. (2002). Psychoanatomical substrates of Balint's syndrome. *J. Neurol. Neurosurg. Psychiatry, 72*(2), 162–178.

Rollenhagen, J. E., & Olson, C. R. (2000). Mirror-image confusion in single neurons of the macaque inferotemporal cortex. *Science, 287*, 1506–1508.

Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron, 27*(2), 205–218.

Saffran, E. M., & Coslett, H. B. (1996). "Attentional dyslexia" in Alzheimer's disease: A case study. *Cogn. Neuropsychol., 13*, 205–228.

Sawamura, H., Georgieva, S., Vogels, R., Vanduffel, W., & Orban, G. A. (2005). Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *J. Neurosci., 25*(17), 4294–4306.

Schmahmann, J. D., & Pandya, D. N. (2006). *Fiber pathways of the brain*. Oxford, UK: Oxford University Press.

Sereno, M. I. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science, 294*, 1350–1354.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends Cogn. Sci., 7*(11), 489–493.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. USA, 104*(15), 6424–6429.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, UK: Cambridge University Press.

Shallice, T., & Warrington, E. K. (1977). The possible role of selective attention in acquired dyslexia. *Neuropsychologia, 15*(1), 31–41.

Sieroff, E., Pollatsek, A., & Posner, M. I. (1988). Recognition of visual letter strings following injury to the posterior visual spatial attention system. *Cogn. Neuropsychol., 5*, 427–449.

Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron, 46*(5), 823–835.

Skarratt, P. A., & Lavidor, M. (2006). Magnetic stimulation of the left visual cortex impairs expert word recognition. *J. Cogn. Neurosci., 18*(10), 1749–1758.

Somers, D. C., Dale, A. M., Seiffert, A. E., & Tootell, R. B. H. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc. Natl. Acad. Sci. USA, 96*, 1663–1668.

Suzuki, K., Yamadori, A., Endo, K., Fujii, T., Ezura, M., & Takahashi, A. (1998). Dissociation of letter and picture naming resulting from callosal disconnection. *Neurology, 51*, 1390–1394.

Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain, 122*(Pt. 11), 2119–2132.

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Liu, A. K., & Dale, A. M. (1998). The representation of the ipsilateral visual field in human cerebral cortex. *Proc. Natl. Acad. Sci. USA, 95*(3), 818–824.

Treisman, A., & Souther, J. (1986). Illusory words: The roles of attention and of top-down constraints in conjoining letters to form words. *J. Exp. Psychol. Hum. Percept. Perform., 12*(1), 3–17.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends Cogn. Sci., 11*(2), 58–64.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Valdois, S., Bosse, M. L., & Tainturier, M. J. (2004). The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attentional disorder. *Dyslexia, 10*(4), 339–363.

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron, 43*(2), 271–282.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron, 55*(1), 143–156.

Vinckier, F., Naccache, L., Papeix, C., Forget, J., Hahn-Barma, V., Dehaene, S., et al. (2006). "What" and "Where" in word reading: Ventral coding of written words revealed by parietal atrophy. *J. Cogn. Neurosci., 18*, 1998–2012.

Warrington, E. K., Cipolotti, L., & McNeil, J. (1993). Attentional dyslexia: A single case study. *Neuropsychologia, 34*, 871–885.

Warrington, E. K., & Shallice, T. (1980). Word-form dyslexia. *Brain, 103*(1), 99–112.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Q. J. Exp. Psychol. [A], 50*, 439–456.

Whiting, W. L., Madden, D. J., Langley, L. K., Denny, L. L., Turkington, T. G., Provenzale, J. M., et al. (2003). Lexical and sublexical components of age-related changes in neural activation during visual word identification. *J. Cogn. Neurosci., 15*(3), 475–487.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychon. Bull. Rev., 8*(2), 221–243.

Whitney, C., & Lavidor, M. (2004). Why word length only matters in the left visual field. *Neuropsychologia, 42*(12), 1680–1688.

Wright, N. D., Mechelli, A., Noppeney, U., Veltman, D. J., Rombouts, S. A., Glensman, J., et al. (2007). Selective activation around the left occipito-temporal sulcus for words relative to pictures: Individual variability or false positives? *Hum. Brain Mapp, 29*(8), 986–1000.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychol. Bull., 131*(1), 3–29.

Zihl, J. (1995). Eye movement patterns in hemianopic dyslexia. *Brain, 118*(Pt. 4), 891–912.

55 The Neural Basis of Syntactic Processing

DAVID CAPLAN

ABSTRACT Syntactic structures are unique mental representations that relate the meanings of words to one another. Understanding of the neural basis for syntax consists mostly of information about the areas in which these structures are assigned and used to determine meaning in the process of comprehension, together with electrophysiological correlates of these processes. This chapter briefly reviews deficit-lesion correlations and neurovascular studies that are relevant to the first of these topics. Both these sources of data suggest that the brain does not support syntactic processing in an abstract fashion but as part of performing the task that is the purpose of the comprehension process and that these task-related syntactic operations are supported by multiple brain areas.

Marr (1982) articulated a useful framework for describing cognitive functions. In this system, a cognitive function is described at three levels: a level at which the representations of the information in the cognitive domain are described (the representational level), a level at which the operations that compute these representations are described (the algorithmic level), and a level at which the neural mechanisms that support the storage of the representations and the activity of the operations that compute them are described (the neural level). In this chapter, I briefly review syntactic processing using this framework to organize the presentation. Readers may find that more space is devoted to the representational and algorithmic levels than is the case in other chapters. If so, this emphasis reflects my sense that these levels are less well understood by neurologically oriented cognitive neuroscientists in this domain than may be the case in other cognitive areas.

Syntactic representations and their processing

Sentences convey information beyond that which is conveyed by words alone. This information, collectively known as the propositional content of a sentence, includes who is initiating and receiving an action (thematic roles), which adjectives are assigned to which nouns (attribution of modification), which words refer to the same items (co-reference), and other similar information, mostly relating to the relationships between items, actions, and properties referred to by the words in a sentence. Propositions are the source of much of the information that is stored in semantic memory. In addition, because propositions can be true or false, they can be used to reason, including making inferences, and to plan actions. Without propositions, language would consist of designating items, actions, and properties of items and actions—a significant functional capacity, to be sure, but far less rich and useful than that which language affords because it includes propositions.

For sequences of words to convey propositional relationships in a flexible manner—one that allows unlikely or impossible relationships to be expressed—it is necessary that the meanings of words be combinable into propositions in some way that does *not* correspond to likely events. That is, combinatory possibilities have to be available to allow the sequence of words "man dog bite" to be associated with the proposition that a man is biting a dog, and not vice versa. Humans use the ability to refer to unlikely and false events when they lie, when they consider hypothetical situations, and in other circumstances. The principles that allow these functions are the syntactic structures of language.

Syntactic structures need not be complex to permit unlikely propositions to be expressed: a simple active form ("The man is biting the dog") would suffice for this basic purpose. But syntactic structures are much more complex than this one requirement imposes, and the complexity adds to the semantic information they allow language to convey. Features of syntax such as embedding allow *propositions*, not merely *words*, to be related to one another. The sentence "The man who chased the girl fell down" expresses a relationship between two propositions—the man chased the girl, and the man fell down. The syntactic structure known as a relative clause allows these two propositions to apply to the same man. Similarly, complement structures allow us to express propositional attitudes: "John believed/disagreed/expected/feared that it would rain" expresses a variety of states of mind that *John* is in vis-á-vis the proposition that it will rain. Syntactic structures are needed to allow these sorts of relations between propositions to be conveyed.

DAVID CAPLAN Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts

In addition to propositional information, syntactic structures also express discourse-level information. The sentences "It was the policeman who shot the robber" and "It was the robber who the policeman shot" convey the same thematic roles, but the first sentence makes the policeman the topic and the second makes the robber the topic. To convey *both* propositional and discourse information, it must be possible to express the same proposition in different forms—whence another functional role for a complex syntax.

The most widely accepted theories of syntactic structures view them as hierarchically organized sets of syntactic categories, over which configurational relations are defined that determine these propositional aspects of meaning (figure 55.1). Different syntactic relations determine different aspects of meaning. In "The boy who chased the dog fell," *the dog* has no semantically consequential syntactic relation to *fell*. In the sentence "The boy who chased the dog caught him," the very similar (though not identical) relation of *the dog* to *him* allows *the dog* to serve as the antecedent of *him*. Similarly, *the boy* is the syntactic subject of both sentences and plays a thematic role around the verb *fell* in both, but cannot be the antecedent of *him* in "The boy who chased the dog caught him."

Basic questions in linguistics revolve around the details of these structures. In the model developed by Chomsky (1995), for instance, syntactic structures are formed by merging lexical categories into higher-level phrasal categories, copying many categories into grammatical positions that need to be filled, using the resulting structure to determine meaning, and deleting the categories in their positions of origin to determine the phonological form of the sentence. Other theories (e.g., Goldberg, 1995, 2006) do not postulate "underlying" syntactic structures: the structure that is "visible" on the surface is the only one generated.

It is beyond the scope of this essay to discuss the differences between these theories in detail, but it is important to provide the reader with some idea of the sorts of issues involved in deciding between alternative conceptions of syntactic structures. As an oversimplified introduction, consider

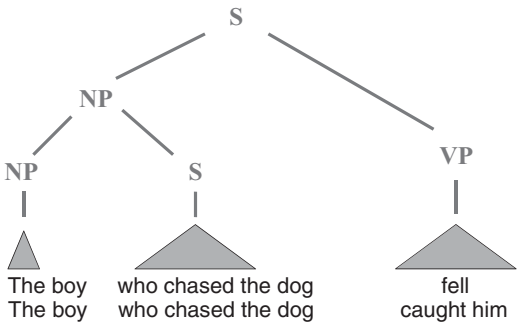


FIGURE 55.1 Schematic representation of hierarchical syntactic structure.

the important issue of whether the surface form of a sentence completely captures its syntactic structure. Chomsky has long argued that such models miss important generalizations about how syntactic relations are related to meaning. For instance, in general, verbs assign a thematic role to their objects, but there are many sentences in which this is not the case when we consider the surface syntactic form—for example, sentences with object-extracted relative clauses ("The boy who the dog chased fell"), questions ("Who did the dog chase?"), indirect questions ("The boy knew who the dog chased"), passives ("The boy was chased by the dog"), and other structures. In all these cases, Chomsky postulates that *the boy* (or *who*) appears in the position of the object of the verb *chase* at some level of syntactic structure and moves to its surface position, leaving a copy of its former self to which the thematic role is assigned. However, Goldberg and many others have argued that the surface forms of sentences introduce particular semantic properties that are not found in a plausible "underlying" syntactic form. For instance, many linguists in the Chomskian tradition have argued that the "ditransitive" construction (e.g., "John tossed Mary a bottle opener") is derived from a "dative" structure (e.g., "John tossed a bottle opener to Mary") in which the thematic roles of agent, theme, and goal follow the more common ("canonical") English linear order. Goldberg, however, points out that the ditransitive has semantic properties not found in the dative—it requires an animate recipient, for instance: one cannot say, "John tossed the counter a bottle opener", but "John tossed a bottle opener to the counter" is perfectly acceptable)—and thus the correct syntactic theory recognizes surface constructions that are associated with aspects of propositional meaning, not surface structures that are derived from underlying representations. The reader can appreciate that it is not possible to develop detailed models of how syntactic information is represented and processed in the brain without knowing what syntactic representations are, and therefore that these disagreements present challenges to understanding the neural basis for syntactic processing. Models of the neural basis for representing and processing syntactic structure have dealt with this uncertainty by either adopting a particular theory or by investigating phenomena that are thought to be basic to syntactic representations and processing regardless of how they are expressed in a theory.

To conclude this section on syntactic representations, we note that a related, but secondary, question is the extent to which syntactic structures differ from representations in other cognitive domains such as mathematics and music (including "languages" naturally used by and taught to animals). Suffice it to say that, although when considered at a sufficiently abstract level, representations proposed in these other areas share features with syntax, virtually all models of the syntax of natural human languages postulate

structures that differ in fairly important ways from those postulated in these other domains. Certainly the contribution of syntactic relations between words to meaning differs from the contribution of the rules relating elements in these other domains to meaning, because of the differences in the meanings conveyed by sentences and these other representational systems.

Most models of sentence comprehension maintain that "parsing" and "interpretive" operations assign syntactic structures and use them to determine aspects of meaning in the process of understanding a sentence. These models have articulated principles whereby parsing rules apply. For instance, in the sentence "The boy wanted to go to the game yesterday," *yesterday* preferentially modifies *to go* not *wanted*, suggesting a general principle of attaching new phrases to the last incomplete phrase in a developing syntactic structure. Sentence interpretation relies on more than just syntactic structure and word meanings; information about the frequency with which constructions appear, the plausibility of the meaning of a sentence, and other factors affect the ease of syntactic analysis and comprehension (MacDonald, Pearlmutter, & Seidenberg, 1994). For instance, "While the man ate the hot dog burned in the fire" is harder to structure and interpret than "While the man ate the wood burned in the fire," because *the hot dog*, but not *the wood*, is a plausible theme of *ate*, which reinforces an ultimately incorrect structure. The principles affecting structure building and these other factors interact online (as sentences are analyzed syntactically and assigned meaning). Recent studies provide evidence that features of the nonlinguistic environment in which a sentence is uttered (e.g., the nature of items visible to a listener) enter into these interactions (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

This chapter will review studies that provide data relevant to the neural basis of parsing and interpretive operations. Only results of deficit-lesion correlations in patients with focal lesions and neurovascular activation studies in normal subjects will be covered. Some other potential sources of data (intraoperative stimulation, subdural electrode placement, transcranial magnetic stimulation, magnetoencephalography, intraoperative and subdural recordings, optical imaging) have not been extensively used for these studies; for review of electrophysiological studies, see Hagoort, Baggio, and Willems (chapter 56 in this volume). Only studies of comprehension will be reviewed, as there is more work in this area than in production.

Deficit-lesion correlation studies of syntactic processing

The logic underlying the use of deficits to explore the neural basis of syntactic processing is that, if a patient's performance can be analyzed as being due to a deficit in a syntactic operation (plus residual abilities and any compensatory strategies that apply), the deficit is due to the lesion the patient has sustained. The corollary of this statement is that the integrity of the lesioned area/neural process is necessary for the operation to take place. To apply deficit-lesion analyses to the problem of localization of syntactic operations, it is therefore necessary to characterize the deficits in patients, their lesions, and the relations between the two.

Two basic views of deficits affecting syntactically based comprehension have been articulated. The first is that individual parsing or interpretive operations are selectively affected by brain damage. The second is that patients lose the ability to apply what have been called "resources" to the task of assigning and interpreting syntactic structure. The first of these deficits may be likened to a student not being able to calculate π to eight decimal places in his/her head because s/he does not know the formula for calculating π. The second may be likened to a student knowing the formula but not being able to hold the intermediate products of computation in mind. Exactly what prevents the application of such knowledge is unclear, but most models of cognitive processes include limitations of this sort.

The hallmark of a deficit affecting syntactic operations is the combination of abnormally low (or chance) performance in understanding sentences that require a syntactic analysis to be understood—semantically reversible sentences that cannot be understood by the application of simple heuristics such as the assignment of thematic roles to nouns following a simple pattern (e.g., "The boy who the girl pushed is tall")—and the retained ability to understand "semantically irreversible" sentences with the same syntactic structures, that is, sentences in which the meaning can simply be inferred from the meanings of the words and knowledge about likely relations between them (e.g., "The book that the girl read is long") (Caramazza & Zurif, 1976). Researchers who advocate "specific deficit" accounts of aphasic disturbances in this area have claimed that this pattern occurs for representations and processes specified in linguistics and psycholinguistic models. Some of these deficits are said to be very specific. For instance, the "trace deletion hypothesis" (Grodzinsky, 2000) maintains that individual patients cannot process sentences that Chomsky's theory maintains contain a certain type of moved items (the term "trace deletion hypothesis" refers to an earlier version of Chomsky's theory in which these items were moved and left a "trace," not copied). The claim is that some patients have lost the ability to connect certain moved (or, now, copied) noun phrases to their "traces" in sentences such as those mentioned previously (relative clauses, questions, indirect questions, passive), with the consequence that these noun phrases are not assigned thematic roles. At the other end of the spectrum, some researchers have suggested that certain aphasics have deficits that apply to a large set of related operations, such as the operations that map all syntactic structures

onto propositional meaning (the "mapping hypothesis"—Linebarger, Schwartz, & Saffran, 1983).

Evidence for such deficits would come from the finding that a patient had an impairment restricted to processing the sentences that required that structure or operation to be understood. Proponents of these models have argued that there are data from aphasia of this sort, but there are three important limitations to such data.

First, the data usually consist of one measure of performance—usually accuracy. To my knowledge, there are three studies in the literature—Tyler (1985), Caplan and Waters (2003), and Caplan and colleagues (2007a)—in which both accuracy and response-time (RT) data have been reported on the same sentences in aphasic patients. Without both accuracy and RT data, it is impossible to rule out speed-accuracy trade-offs as the source of selective impairments, or to know if a patient's problem is manifest only in a longer time that s/he requires to process a syntactic structure. The same three studies are among the few in which online and end-of-sentence measures have been gathered on the same sentences in the same patients. Online data are critical in many ways. For instance, a patient may show normal online performance and an impairment in performing a task; this finding would suggest that the patient assigns the normal structure and interpretation of a sentence of a certain type but fails to use the meaning s/he extracts normally—because of limitations on how long it can be retained, because alternative interpretations (some of which may be activated in the course of normal comprehension) are not inhibited, or for other reasons. In situations such as these, it would not be correct to say that the patient has a deficit in assigning or interpreting the syntactic structure in question.

Second, patients have usually been tested on only one task that requires comprehension, most often sentence-picture matching. However, it is well documented that performance may dissociate over tasks (Cupples & Inglis, 1993; Caplan, Waters, & Hildebrandt, 1997). Caplan, DeDe, and Michaud (2006) and Caplan and colleagues (2007a) studied 42 aphasic patients and found only two in which the same deficit appeared in sentence-picture matching and object manipulation. An inability to perform accurately on a set of sentences in one comprehension task cannot be taken as a reflection of an impairment of a syntactic operation if the patient can perform accurately on those sentences in another comprehension task.

Third, important linguistic controls have usually not been run that would show that the deficit is restricted to the structures claimed. For instance, the trace deletion hypothesis (Grodzinsky, 2000) maintains that patients with the deficit in question are able to co-index items other than traces, such as pronouns. Thus a patient who has this deficit would not be able to connect *the boy* to the "trace (t)" in "The boy who the man pushed t bumped him," but s/he should be able to connect *the man* to the *him* in this sentence. Accordingly, the patient should not know who pushed whom, but should know that the boy bumped the man. None of the papers in the literature that have been taken as supporting the trace deletion hypothesis have reported patients' performance on both sentences with "traces" and sentences with pronouns (or other referentially dependent items, such as reflexives—see Caplan, 1995; Caplan et al., 2007a, for discussion).

Much of the evidence for specific syntactic deficits is based not upon the performance of individual patients but upon the performance of small groups of patients with certain diagnoses drawn from the traditional clinical literature on aphasia, such as Broca's aphasia or agrammatic aphasia. It is sometimes claimed that some of the objections that we have raised are answered by these group data. For instance, Grodzinsky (2000) has argued that some agrammatic patients have shown integrity of processing pronouns, answering questions about the adequacy of linguistic controls. In the view of this writer, these studies do not address the issues raised. For instance, although some agrammatic patients have shown normal performances on sentences with pronouns (Grodzinsky, Wexler, Chien, Marakovitz, & Solomon, 1993), these patients have not also been tested on sentences with "traces," so we do not know if they show the deficit specified by the "trace deletion hypothesis." Empirical data show that not all patients with a clinical diagnosis of Broca's aphasia or agrammatism have problems with sentences containing "traces" (Swinney & Zurif, 1995; Zurif, Swinney, Prather, Solomon, & Bushell, 1993; Blumstein et al., 1998; see Berndt, Mitchum, & Haendiges, 1996; Drai & Grodzinsky, 1999, Caramazza, Capitani, Rey, & Berndt, 2001; Caplan, 2001a, 2001b, for discussion), so there is a need to verify the presence of both deficits on a patient-by-patient basis.[1]

Turning to the hypothesis that deficits of aphasic syntactic comprehension should be characterized as reductions of processing capacity, four arguments have been made in support of this suggestion: (1) the finding that some patients can understand sentences that contain certain structures or operations in isolation but not sentences that contain combinations of those structures and operations (Caplan & Hildebrandt, 1988; Hildebrandt, Caplan, & Evans, 1987); (2) the finding that, in large groups of patients, as patients' performances deteriorate, more complex sentence types are affected more than less complex ones (Caplan, Baker, & Dehaut, 1985; Caplan et al., 2007a); (3) the fact that, in factor analyses of performance of such patient groups in syntactic comprehension tasks, first factors on which all sentence types load account for the majority of the variance (Caplan et al., 1985, 2007a; Caplan, Hildebrandt, & Makris, 1996); (4) the claim that simulations of the effect of resource reductions on syntactic comprehension in normal subjects

through the use of speeded presentation (Miyake, Carpenter, & Just, 1994), concurrent tasks (King & Just, 1991), and other methods mimic aphasic performance.

These arguments are also not ironclad. The argument that some patients can understand sentences that contain certain structures or operations in isolation but not sentences that contain combinations of those structures and operations suffers from the same limitations of the database that we discussed earlier: it is based on a single performance measure (accuracy) in a single task (enactment). Testing the second result—that as patients' performances deteriorate, more complex sentence types are affected more than less complex ones—risks circularity unless the effects of resource reduction are modeled and measured separately from comprehension on sentences of the sort that are used to test the effects of complexity. Three studies have addressed this issue in different ways: two (Caplan et al., 1985; Caplan et al., 2007a) have found this pattern; the third, a smaller study, did not (Dick et al., 2001). The data regarding interference effects in normal subjects are complex. Interactions of load and syntactic complexity, and of these factors with subject groups that differ in processing resource capacity, are critical pieces of evidence that would support this model, but these interactions only occur under special circumstances (see Caplan & Waters, 1999; Caplan et al., 2006, for reviews), reducing the strength of this argument. The finding that first factors on which all sentence types load account for the majority of the variance is an extremely robust finding, regardless of the task over which factors are extracted or whether they are extracted over several tasks (DeDe Caplan, 2006; Caplan et al., 2007a). The hypothesis that reductions of processing capacity are sources of aphasic syntactic comprehension deficits fares better, in my view, than the hypothesis that individual patients have specific deficits.[2]

Accepting the view that what is to be correlated with lesion parameters is either some measure of performance that captures a deficit a patient has with a particular syntactic structure or operation, or some measure of performance that captures the "amount" of resources available to a patient, what do studies of deficit-lesions correlation show about the way the brain is organized to support parsing and interpretation? Four models of brain organization for syntactic processing have been suggested, based on data of this sort. Localizationist models are represented by Grodzinsky (2000), who claims that Chomskian traces are coindexed in Broca's area; variable localization models by ourselves (Caplan, 1994; Caplan et al., 2007b); invariant evenly distributed models by Dick and colleagues (2001) and Damasio and Damasio (1992); and invariant unevenly distributed models by Mesulam (1990, 1998). Most of these models have been articulated as applying to specific operations, but the evidence can often be interpreted in terms of reductions in the resource system that underlies parsing and interpretation. For instance, unless the proper control studies are done (discussed previously), performances that are interpreted as failures to "co-index traces" can be seen as due to reductions in processing resources that lead to failures to comprehend sentences that contain "traces."

We must begin with a major caveat about lesion-deficit studies: the vast majority of these studies do not examine lesions quantitatively. Many are based on the assumption that Broca's or nonfluent patients have "anterior" lesions and Wernicke's, fluent, conduction, and anomic patients have "posterior" lesions, whereas the reality is far more complicated (Mohr et al., 1978; Vanier & Caplan, 1989). A number of studies summarize radiological reports and/or display lesions, usually on a single transverse section of the brain imaged with computer tomography or magnetic resonance, and emphasize the area in which lesions in patients with certain types of performances (analyzed as deficits of particular types) overlap. Such analyses do not investigate many questions. For instance, neither the most direct prediction made by distributed models (that lesion size correlates with performance level) nor the claim that the insertion of traces into syntactic structures occurs in Broca's area has ever been tested on the basis of radiological data by advocates of these models (Mesulam, 1990; Damasio & Damasio, 1992; Dick et al., 2001; Grodzinsky, 2000).

To my knowledge, there are only six studies in the literature in which radiological images have been analyzed and related to sentence comprehension in aphasics. Three are based on instruments that do not examine syntactic processing: Karbe and colleagues (1989), who used the Western Aphasia Battery, which does not characterize deficits in a linguistically or psycholinguistically specific way; Kempler, Curtiss, Metter, Jackson, and Hanson (1991), who used the Token Test, which confounds syntactic processing with short-term memory requirements; and Dronkers, Wilkin, Van Valin, Redfern, and Jaeger (2004), who used the CYCLE, which does not separate lexical from syntactic errors. Two studies used appropriate measures to test syntactic processing but had other limitations. Tramo, Baynes, and Volpe (1988) presented reversible sentences in a sentence-picture matching task, but studied only one contrast (active and passive sentences) and only reported three cases. Caplan and colleagues (1996) studied 25 sentence types testing many aspects of syntactic processing, but only 18 patients were studied, lesions were identified subjectively, and scans were normalized along a single linear dimension in the anterior-posterior plane only, likely leading to significant inaccuracies in the estimates of percents of regions of interest (ROIs) that were lesioned. Other problems are found in many of the studies that also used inappropriate test instruments (see Caplan et al., 2007b, for discussion). In all these studies, analyses were limited to examining the effect of lesions in individual locations; only Caplan and colleagues

(1996) considered the effect of size of lesion or the relative effects of lesions in multiple regions on performance.

Caplan and colleagues (2007b) began to address these issues. We studied 42 right-handed native English-speaking aphasic patients with single left-hemisphere strokes and 25 age- and education-matched controls. Patients were tested in object manipulation, sentence-picture matching, and grammaticality judgment tasks, the latter two with whole-sentence auditory and word-by-word self-paced auditory presentation ("auditory moving windows"; Ferreira, Henderson, Anes, Weeks, & McFarlane, 1996). In each task, three syntactic operations were tested: passivization, object relativization, and co-indexation of a reflexive. First-factor scores of principal components analyses of performance on each task were taken as reflections of resource availability of each patient in each task, and differences in accuracy, response time, and listening times for words in critical positions (corrected for word length and frequency) in experimental and baseline sentences served as measures of syntactic processing ability for particular structures.

Thirty-two patients and 13 controls matched for age and education underwent magnetic resonance (MR) and positron emission tomography (PET) scanning. The relation between lesions and deficits was investigated by regressing MR and PET measures of the extent of the lesion in each of seven ROIs against the measures of syntactic processing mentioned earlier. The results showed that percent lesion volume on MR and mean PET counts/voxel in several small regions accounted for a significant amount of variance in performance measures. For instance, percent MR lesion in the inferior parietal lobe, the anterior inferior temporal lobe, and the superior parietal lobe, and PET counts/voxel in Broca's area accounted for a significant amount of variance in first-factor scores for all tasks combined and for object manipulation; different patterns of predictor variables were found for other dependent variables. Since the number of cases was small, a simpler analytic approach was also used: the range of performance in patients with lesions within 0.25 standard deviation of the mean lesion size in four regions (the entire left hemisphere, the left hemisphere cortex, the perisylvian association cortex, and the combination of the perisylvian association cortex, the inferior anterior temporal lobe, and the superior parietal lobe) was examined. In each case, performances covered a wide range of total performance; in some cases almost the entire range, and in one case the entire range, of performance was found.

These results provide strong evidence against models that maintain that the operations or resources involved in assigning and interpreting specific sentence types are invariantly localized in one brain area. If this were the case, measures of lesion extent in only one area should have predicted the magnitude of a deficit. Similarly, these results argue against models that maintain that these functions are evenly distributed across large contiguous brain areas, such as the left-hemisphere cortex or the perisylvian association cortex. If this were the case, lesions of equal size in areas such as those considered should have led to similar magnitudes of a deficit. The data are consistent with the view that localization of these operations and resources varies across individuals. They are also consistent with the idea that the functions being measured are unevenly distributed throughout sets of apparently otherwise unrelated cortical areas, either with the same pattern of unevenness in all individuals ("invariant uneven distribution") or with different patterns of uneven distribution in different individuals ("variable uneven distribution"). The evidence that would discriminate between these models is not available (see Caplan et al., 2007b, for discussion).

An important finding in this study is that the patterns of significant effects of lesion sites differed for different tasks. This suggests that lesions cause deficits in operations or resource systems that are deployed in certain tasks, not operations or resource systems that are used to assign and interpret syntactic structures in an amodal, abstract, fashion. This conclusion is consistent with the finding of Caplan and colleagues (2006) that deficits affecting the same operation on two tasks are very rare. They suggest that the brain is organized such that regions support syntactic operations in particular tasks.

Functional neuroimaging studies of syntactic processing

Functional neuroimaging has supplanted deficit-lesion correlations as the principal source of information regarding the neural organization that supports syntactic processing in sentence comprehension. Functional neuroimaging in intact individuals provides different information about functional neuroanatomy than deficit-lesion correlations, namely, information about (a subset of) the neural areas and processes that are normally sufficient to support a function.

As with deficit-lesion correlations, functional neuroimaging studies can be divided into those that investigate specific operations and those that examine the "processing resource" system that supports these functions. We will review these areas in turn (for more extensive reviews of this literature see Grodzinsky & Freiderici, 2006; Caplan, 2006, 2007).

In a line of research that parallels his work in aphasia, Grodzinsky and his colleagues have suggested that fMRI studies implicate left posterior inferior frontal gyrus (Broca's area) as the site responsible for co-indexing traces. Ben-Shachar, Hendler, Kahn, Ben-Bashat, and Grodzinsky (2003) contrasted object-relativized sentences (sentence 1), in which this operation occurs, with complements (sentence 2), in which it does not, in a grammaticality judgment task ([t] = trace; identical subscripts indicate referential co-indexation).

1. I helped the girl$_i$ that Mary saw [t$_i$] in the park.

2. I told Mary that the girl ran in the park.

BOLD signal increased in left inferior frontal gyrus (IFG) and bilateral superior temporal sulcus (STS) in sentence 1 compared to sentence 2. Ben Shachar, Palti, and Grodzinsky (2004) found increased BOLD signal in the left IFG, in the left ventral precentral sulcus, and bilaterally in superior temporal gyrus (STG) (marginally on the right) in the contrast of embedded *wh*-questions (sentences 3 and 4) against yes/no questions (sentence 5) in a verification task:

3. The waiter asked which tourist$_i$ [t$_i$] ordered the alcoholic drink in the morning.

4. The waiter asked which alcoholic drink$_i$ the tourist ordered [t$_i$] in the morning.

5. The waiter asked if the tourist ordered the alcoholic drink in the morning.

These studies are the only studies to date that contrast sentences with and without the co-indexation of a "trace," and they yield activation in more than one region.

Bornkessel and Schlesewsky (2006) have published a major position paper on a neurologically based model of parsing and sentence interpretation. They argue that the linear order of noun phrases is mapped onto thematic roles in Broca's area and that deviations from the usual mapping lead to increased activity in this area. The mapping is determined by both general features of language and cognition (e.g., animate nouns are more likely to be agents than inanimate nouns) and language-specific grammatical features (e.g., case marking is more important in determining thematic roles in a language that has a great deal of visible case marking, such as German, than in a language that does not, such as English). The model is groundbreaking in that it is the first detailed model of aspects of sentence comprehension that is based primarily upon neurological data, mostly event-related potential studies. It is worth noting, however, that the model only deals with a small aspect of sentence processing—assigning the two most basic thematic roles (prototypical agents and themes) in the simplest syntactic structures (single sentences). As noted, the model mostly deals with ERP data, but the anatomical hypotheses are partially based on fMRI studies; I shall briefly review these here.

As noted, the model focuses on the role of Broca's area in mapping noun phrases onto thematic roles. The first critical study of this topic was Bornkessel, Zysset, Friederici, von Cramon, and Schlesewsky (2005), in which participants verified the meaning of German sentences with verb-final complement clauses in which word order, morphological case ambiguity, and verb class (transitive/dative object experiencer) were varied. Bornkessel and colleagues found that, for sentences with subject-before-object word order, sentences with dative-object-experiencer verbs produced greater BOLD signal in left pars opercularis than sentences with transitive verbs and vice versa for sentences with object-before-subject word order. This finding is consistent with the conclusion that this part of Broca's area is involved in mapping the linear order of thematic roles onto the hierarchy of thematic roles. However, the picture is quite complicated. Grewe and colleagues (2005) found that an increase in BOLD signal in left IFG associated with a less common word order (object before subject) was not present when the first noun phrase was a pronoun. Since pronouns preferentially occur in first position in the German middle field, the authors interpreted this result as evidence that particular language-specific syntactic features override the usual effect of word order. However, the result could also indicate that the object-before-subject word order does not always lead to activation in left IFG. Grewe and colleagues (2007) also failed to find that noun animacy had the effects that the theory predicts: an increase in BOLD signal that occured with object-before-subject order was not greater when the object was inanimate. This finding contradicts the hypothesis in Grewe and colleagues (2006) and Bornkessel and Schlesewsky (2006) that left IFG is activated by sentences that require mapping noun phrases (NPs) that violate the animacy principle onto thematic roles. Another issue is that, as in the studies of BOLD signal associated with processing "traces," brain areas other than left IFG have been activated in these studies. In the Bornkessel and colleagues (2005) study, the left STS and inferior parietal sulcus (IPS) showed the same pattern as left IFG, for morphologically unambiguous sentences. Thus the hypothesis that left IFG supports a general function of "decoding the prominence relations between arguments," and that it is the only brain area to do so, is not well established.

The processing resource system underlying parsing and interpretation has been studied in functional neuroimaging studies that examine the difference between object- and subject-extracted structures (e.g., "The boy who the girl chased fell"; "The boy who chased the girl fell"). Object extraction is more demanding than subject extraction and is thought to require more "resources" than subject extraction, for many reasons. These studies use a variety of sentences—cleft sentences, relative clauses, conjoined sentences, *wh*-questions, complement clauses, main clauses, topicalization, and dative shifts in English, German, Dutch, Japanese, and Hebrew. There is great variability in the areas activated in these studies. In Just, Carpenter, Keller, Eddy, and Thulborn (1996), this contrast activated frontal and temporal perisylvian cortex bilaterally. In Stromswold, Caplan, Alpert, and Rauch (1996), left IFG was activated. Cooke and colleagues (2001) found increased BOLD signal in bilateral inferior temporal lobe in this contrast. In other studies of ours (Caplan, Alpert, & Waters, 1998, 1999; Waters, Caplan, Stanzcak, & Alpert, 2003), activation was seen variably in

medial anterior structures (cingulate, middle frontal, and superior frontal gyri) and other areas (right IFG, left thalamus, left superior parietal lobe). There was no effect of the object/subject-extraction contrast in studies by Fiebach, Vos, and Friederici (2004), Fiebach, Schlesewsky, and Lohmann (2005), Indefrey, Hagoort, Herzog, Seitz, and Brown (2001), or Ben-Shachar and colleagues (2003).

Fiebach and colleagues (2004, 2005) considered several reasons for this variability: differences in sensitivity of imaging technology, tasks, the particular structures in which extraction occurred (indirect questions; relative clauses), languages, and subjects (high and low span). To these we can add differences in presentation modality, design (blocked versus event related), normalization methods, modeling of the hemodynamic response, and how the significance of activation was determined (fixed versus random-effects models; omnibus analyses versus preselected ROIs). There is an important point about these factors, however: none of them other than the use of inadequately powerful statistical methods could reveal BOLD signal effects that are not present in a contrast, and therefore, unless many of these studies have used inadequate statistical methods and are reporting false positive effects, there are a large number of areas that are activated by these contrasts, suggesting some type of nonlocalizationist model for the neural structures that provide the resources utilized in parsing and interpreting these sentences. However, two factors that we have not mentioned thus far could account for much of this variation and potentially salvage a localization model. The first is the possibility that ancillary cognitive operations ("strategies") co-occur with parsing and interpretation (Page, 2006) and are the source of some of the BOLD signal effect found in these studies. The second are task-stimulus interactions. We will consider these factors briefly.

First, strategies may have affected neurovascular effects. We investigated this possibility in a verification task (Caplan, Chen, & Waters, 2008). We presented sentences for 4 seconds, followed by a fixation point for 2 seconds, followed by a probe in an active form. We separated BOLD signal into an "early" set of timed repetition (TR) intervals associated with processing the target and a "late" set associated with processing of the probe. In the early TR intervals, there were a few areas of paradoxically greater BOLD signal for syntactically simple compared to complex sentences. This result may have occurred because participants responded to the easier sentences more quickly. In later TR intervals, BOLD signal increased in four left perisylvian locations—the inferior frontal gyrus, the middle frontal gyrus, the inferior parietal sulcus, and the middle temporal gyrus—and in a variety of other areas in response to the complex sentences. This finding strongly suggests that many areas of activation in this task are due to retaining syntactic representations or a representation from which syntactic representations were constructed in memory. Some of the activity associated with sentence contrasts in the literature cited previously may thus be a result of retaining these representations, not constructing them. Strategies of this sort may have led to BOLD signal effects associated with sentence contrasts in other studies, leaving open the possibility that the resource system that supports parsing and interpretation itself is localized in one area.

Second, performing a task affects parsing and interpreting specific sentences, and the interaction of performing a task and parsing and interpretation may result in the BOLD signal associated with a sentence contrast. Let us take plausibility judgment as an example. There is strong evidence that subjects assess the plausibility of thematic assignments incrementally as syntactic constituents are constructed and interpreted (e.g., Trueswell, Tanenhaus, & Garnsey, 1994; Pearlmutter & MacDonald, 1995). These assessments of plausibility can be used incrementally to weight responses in a plausibility judgment task (Garnsey, Tanenhaus, & Chapman, 1989; Boland, Tanenhaus, Garnsey, & Carlson, 1995). There is a complete isomorphism between the assignment of thematic roles and this function. For example, more thematic roles are assigned at the embedded verb of object-extracted sentences than at any point in subject-extracted sentences, and therefore more plausibility judgment weighting can occur there as well. BOLD signal differences between sentences might thus reflect different processing loads associated with incremental plausibility judgment in different sentences, not different demands of task-independent parsing and interpretation of different sentence types.

To see if this might be the case, we compared BOLD signal responses to the same sentences in two tasks: plausibility judgment and nonword detection (Caplan et al., 2008). The plausibility-judgment task led to widespread activation in the contrast of syntactically complex and simple sentences. In nonword detection, participants viewed the same plausible sentences containing only real words as they saw in the plausibility-judgment task, and an equal number of sentences in which a word had been replaced by an orthographically and phonologically legal nonword. Nonword detection showed a very different pattern of BOLD signal effects. For sentences containing only real words—the same sentences as the plausible sentences that were analyzed in the plausibility-judgment task—there was an increase in BOLD signal located only in left BA 44. The behavioral data showed that subjects not only searched each stimulus for nonwords but also processed the sentences as sentences. The results thus suggest that, in plausibility judgment, the BOLD signal effects largely resulted from the incremental weighting of the response selection mechanism by plausibility information that was activated online. In contrast, left BA 44 may be the area in which task-independent operations that are used in the more complex but not the simpler sentences are localized.

These considerations suggest that, once the effects of strategies and task-sentence type interactions are eliminated, specific parsing and interpretation operations may yet be supported by a limited number of brain areas, perhaps only one. Other data show that the picture is more complicated, however (Caplan & Waters, 2007). We used the same correct sentences used in the plausibility and nonword-detection tasks, containing only real words, in a third task—font-change detection. Participants saw these sentences and foils consisting of grammatically correct, meaningful sentences containing only real words, one of which appeared in a slightly different font from the others, and were required to indicate whether a sentence had a font change. Analyses of the behavioral data again showed that subjects processed sentences as sentences. For sentences containing only words without font changes—the same grammatical, plausible sentences that were analyzed in the plausibility-judgment task and in the nonword-detection task—there was an increase in BOLD signal, but it was located in the left supramarginal gyrus (left BA 39), not left IFG. There were no areas activated in both nonword detection and font-change detection, and no functional connectivity between the areas activated in the two tasks. These results indicate that different parsing and interpretive operations were applied in the two tasks—a result that was confirmed by the finding that the effects of the position of a nonword or a word with a font change on detection response times differed. Thus, although individuals do assign and interpret syntactic structures even when these structures are completely irrelevant to task performance, the task they are performing still affects which operations they deploy. To identify the neural basis of particular parsing operations thus requires knowing what parsing and interpretive operations are applied in a task, as well as knowing that neurovascular effects are not due to strategies or task-sentence type interactions.

Concluding comments

The past 15 years have seen great changes in models of parsing and sentence interpretation. For close to three decades (roughly 1965–1995), heavily influenced by Chomsky's views regarding the domain specificity of syntactic representations and Fodor's (1972) concept of modular cognitive processes, researchers studying syntactic processing made the assumption that parsing and interpretive operations were task independent, and that the use of the products of the interpretive process to perform tasks occurred independently of the assignment of syntactic structure and propositional meaning. Correspondingly, deficit-lesion correlations were interpreted as providing evidence for the location of neural tissue that supports task-independent syntactic operations. Most functional neuroimaging studies of syntactic processing have been interpreted within the same framework.

In the past decade or so, evidence has accrued that this view is inaccurate in important respects. It may be that there are task-independent parsing and interpretive operations, but there is very strong evidence that the assignment of syntactic structure interacts at the earliest possible moment with other types of information in the process of assigning sentence structure and meaning, such as the assessment of how plausible certain meanings are, the activation of meanings based upon nongrammatical heuristics, the assignment of structure and meaning based upon the frequency of occurrence of particular constructions or sequences of words, and so on (MacDonald et al., 1994). Though all these operations could be regarded as part of a larger, integrated process that assigns sentence meaning, the problem still remains of isolating the operations that assign the grammatically licensed syntactic structure of a sentence and use it to determine the meaning of the sentence (recall that it is this structure that allows sentences to convey unlikely information, to express complex relations between items and propositions, and to convey both propositional and discourse-level information). Even more unexpectedly from a "modular" point of view, task demands appear to influence parsing and interpretation online; for instance, how one attaches the prepositional phrase in a sentence such as "Put the toy on the rug . . ." depends upon how many toys are in an array that is being inspected and where these toys are located (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The application of parsing and interpretive operations thus may differ in different tasks, and, more seriously from the point of view of identifying the neural basis of these operations, differences in how a task is performed as a function of the sentences that are being presented may be responsible for differences in neural activity associated with sentence contrasts and for deficits that affect performance on one sentence type in a given task. Finally, many tasks involve strategic use of cognitive operations, such as subvocal rehearsal, that are applied to a greater extent when an individual is presented with more complex sentences. These ancillary cognitive operations must also be eliminated from consideration if the neural basis of parsing and interpretation is to be identified. The results of recent studies of patients and neurovascular responses to syntactic contrasts have led to some findings that suggest that these questions are important to consider.

With respect to task dependency of deficits and activation, recent lesion studies have shown that deficits affecting particular parsing operations or the resource system that supports them are affected by task, and the same is true of activation associated with sentence contrasts. These results indicate that most of the data obtained thus far regarding parsing and interpretation may identify brain regions that support a combination of parsing and interpretation and performance of particular tasks. Areas of the brain that are always activated by a syntactic contrast regardless of task are

candidates for the neural substrate of parsing and interpretive operations that are applied regardless of task. Though there are hints in the literature that such areas exist, evidence of this sort is sparse. Two examples that we have reviewed are evidence that the operations that relate the prominence of a noun phrase to its thematic role and those that relate a "moved" constituent to its underlying position invariably involve left IFG. I have argued that the evidence for such invariant activation is at best suggestive.

A feature of recent lesion studies is that they have shown that deficits are associated with lesions in multiple unrelated brain areas and are not predicted by the size of lesions in larger brain areas that are reasonably thought to be possible substrates for parsing and interpretation. Similarly, the activation studies that provide the best evidence for invariant activation of an area across tasks have activated several brain areas. These findings suggest an unorthodox view of the way syntactic operations or the resource system that supports them are related to the brain. The standard view is that a given elementary cognitive operation is invariantly distributed over an area of the brain, either large or small. The results cited earlier are not consistent with this model—they are consistent with the view that the functions identified in these studies are localized in different brain regions in different individuals or distributed across diverse brain areas. I shall end this chapter with a comment on the implications of this possibility.

Properties of neurons such as cytoarchitectonics, receptotonics, connectivity, and other genetically determined features determine the computational capacities of an area of the brain. Invariant localization of a cognitive operation would result from these computational capacities being determined by the *specific* neurological features of particular brain areas. In contrast, variable localization or distribution of a function across diverse brain areas would result from these computational capacities being determined by the neurological features that are *common* to several brain areas, coupled with some factor that determines which of the brain areas that have these features supports a function. Suppose, for instance, that any part of the six-layer association isocortex that is connected to primary auditory koniocortex at a synaptic distance of three or less is capable of supporting parsing operations, and that which area actually supports a particular operation depends upon the history of exposure to particular operations. This capability would surely lead to variable localization of such operations or to a state of affairs in which such operations were supported by multiple, possibly otherwise unrelated, areas of cortex.

These possibilities significantly complicate the picture that needs to be considered regarding the neural organization for syntactic processing (and possibly other aspects of language), but such complication is not necessarily a bad thing. Syntactic operations are among the most abstract naturally occurring species-specific operations that develop in all normal humans, and we would be wise to consider a broad range of possibilities regarding their neural basis. An encouraging feature of current work is that the tools to gather relevant data and to develop models are greatly expanded compared to only a few years ago. We shall see where their use leads.

NOTES

1. In addition to this issue, the reliance on clinical syndromes to identify patients with particular processing deficits runs into other problems. One is that most syndromes are very poorly defined, often referring to patients with "relatively intact" performance on one very general type of task (e.g., "comprehension") and "relatively poor" or "impaired" performance on another (e.g., "speech production"); this problem leads to inconclusive debates about which patients have the deficit and fall within the syndrome. Another is that, in many instances, a clinical syndrome is defined by performances that have no obvious connection to the deficit being proposed. This is the case for the "trace deletion hypothesis," which postulates a very specific comprehension disturbance in patients who have particular problems in sentence production (roughly, nonfluent speech, with omission of grammatical elements). In such cases, the relation between the production and comprehension problems needs to be specified. One possibility is that the connection is functional: that the deficit in one task leads to the deficit in the other. In many cases, such as the "trace deletion hypothesis," this connection has not been suggested; it has not been established in any case in which the classic syndromes are the basis for identification of patients in which a deficit in syntactic comprehension is hypothesized to occur. The alternative is that a lesion that produces one deficit also produces the other. A great deal of confusion has been generated by not distinguishing these two ways that syndromes can be related to hypotheses about specific deficits.
2. An important question is, What, exactly, are "resources"? Seen in the most general terms possible, resources are features of (a model of) a cognitive system that allow certain operations to occur and set limits on their occurrence, but do not themselves enter into computations and are not representations. An example would be the presence and number of hidden units in a Boltzmann machine, whose existence extends the computational power of one-level "perceptrons" and whose number affects the types of generalizations that the system achieves. Cognitive functions that might provide resources for parsing and interpretation include working memory (Miyake et al., 1994), phonological short-term memory (Baddeley, 1986; Caramazza, Basili, Koller, & Berndt, 1980; but see Caplan & Waters, 1990), factors that determine speed of processing (rates of activation and decay: Haarmann & Kolk, 1991, 1994; Haarmann, Just, & Carpenter, 1997), factors that affect weights in connectionist systems (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), and factors that affect the efficiency of lexical processing, which can affect syntactic processing (see Caplan & Waters, 1990, for discussion). All these factors have been considered as a resource in which reduction affects parsing and

interpretation, but none have been definitely shown to play this role. Deficits in a short-term semantic memory system are thought to lead to quite specific, different disturbances in comprehension (Martin & He, 2004).

REFERENCES

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D., & Grodzinsky, Y. (2003). The neural reality of syntactic transformations: Evidence from fMRI. *Psychol. Sci.*, *14*, 433–440.

Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). The neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *Neuroimage*, *21*, 1320–1336.

Berndt, R., Mitchum, C., & Haendiges, A. (1996). Comprehension of reversible sentences in "agrammatism": A meta-analysis. *Cognition*, *58*, 289–308.

Blumstein, S., Byma, G., Kurowski, K., Hourihan, J., Brown, T., & Hutchinson, A. (1998). On-line processing of filler-gap constructions in aphasia. *Brain Lang.*, *61*(2), 149–169.

Boland, J. E., Tanenhaus, M. K., Garnsey, S., M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from *wh*-questions. *J. Mem. Lang.*, *34*, 774–806.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychol. Rev.*, *113*, 787–821.

Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y., & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage*, *26*, 221–233.

Caplan, D. (1994). The cognitive neuroscience of syntactic processing. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 871–879). Cambridge, MA: MIT Press.

Caplan, D. (1995). Issues arising in contemporary studies of disorders of syntactic processing in sentence comprehension in agrammatic patients. *Brain Lang.*, *50*, 325–338.

Caplan, D. (2001a) The measurement of chance performance in aphasia, with specific reference to the comprehension of semantically reversible passive sentences: A note on issues raised by Caramazza, Capitani, Rey and Berndt (2000) and Drai, Grodzinsky and Zurif (2000). *Brain Lang.*, *76*, 193–201.

Caplan, D. (2001b). Points regarding the functional neuroanatomy of syntactic processing: A response to Zurif (2001). *Brain Lang.*, *79*, 329–332.

Caplan, D. (2006). fMRI studies of syntactic processing. *Curr. Med. Imaging Rev.*, *2*, 443–451.

Caplan, D. (2007). Functional neuroimaging studies of syntactic processing in sentence comprehension: A critical selective review. *Lang. Linguistics Compass*, *1*, 32–47.

Caplan, D., Alpert, N., & Waters, G. (1998). Effects of syntactic structure and propositional number of patterns of regional cerebral blood flow. *J. Cogn. Neurosci.*, *10*, 541–552.

Caplan, D., Alpert, N., & Waters, G. (1999). PET studies of sentence processing with auditory sentence presentation. *NeuroImage*, *9*, 343–351.

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, *21*, 117–175.

Caplan, D., Chen, E., & Waters, G. (2008). Task-dependent and task-independent neurovascular responses to syntactic processing. *Cortex*, *44*(3), 257–275.

Caplan, D., DeDe, G., & Michaud, J. (2006). Task-independent and task-specific syntactic deficits in aphasic comprehension. *Aphasiology*, *20*, 893–920.

Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension*. Cambridge, MA: MIT Press (Bradford Books).

Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain*, *119*, 933–949.

Caplan, D., & Waters, G. (1990). Short-term memory and language comprehension: A critical review of the neuropsychological literature. In T. Shallice & G. Vallar (Eds.), *The neuropsychology of short-term memory* (pp. 337–389). Cambridge, UK: Cambridge University Press.

Caplan, D., & Waters, G. S. (1999). Verbal working memory capacity and language comprehension. *Behav. Brain Sci.*, *22*, 114–126.

Caplan, D., & Waters, G. S. (2003). On-line syntactic processing in aphasia: Studies with auditory moving windows presentation. *Brain Lang.*, *84*(2), 222–249.

Caplan, D., & Waters, G. (2007). *BOLD signal response to implicit syntactic processing*. Long Beach, CA: Psychonomic Society.

Caplan, D., Waters, G., & Hildebrandt, N. (1997). Syntactic determinants of sentence comprehension in aphasic patients in sentence-picture matching and enactment tasks. *J. Speech Hear. Res.*, *40*, 542–555.

Caplan, D. Waters, G., Kennedy, D. Alpert, A., Makris, N., DeDe, G., Michaud, J., & Reddy, A. (2007a). A study of syntactic processing in aphasia. I. Psycholinguistic aspects. *Brain Lang.*, *101*, 103–150.

Caplan, D., Waters, G., Kennedy, D., Alpert, A., Makris, N., DeDe, G., Michaud, J., & Reddy, A. (2007b). A study of syntactic processing in aphasia II. Neurological aspects. *Brain Lang.*, *101*, 151–177.

Caramazza, A., Basili, A. G., Koller, J. J., & Berndt, R. S. (1980). An investigation of repetition and language processing in a case of conduction aphasia. *Brain Lang.*, 14, 235–271.

Caramazza, A., Capitani, E., Rey, A., & Berndt, R. S. (2001). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. *Brain Lang.*, *76*, 158–184.

Caramazza, A., & Zurif, E. R. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain Lang.*, *3*, 572–582.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Cooke, A., Zurif, E. B., DeVita, C., Alsop, D., Koenig, P., Detre, J., Gee, J., Pinãngo, M., Balogh, J., & Grossman, M. (2001). Neural basis for sentence comprehension: Grammatical and short-term memory components. *Hum. Brain Mapp.*, *15*, 80–94.

Cupples, L., & Inglis, A. L. (1993). When task demands induce "asyntactic" comprehension: A study of sentence interpretation in aphasia. *Cogn. Neuropsychol.*, *10*, 201–234.

Damasio, A. R., & Damasio, H. (1992). Brain and language. *Sci. Am.*, September, 89–95.

DeDe, G., & Caplan, D. (2006). Factor analysis of syntactic deficits in aphasic comprehension. *Aphasiology*, *20*, 123–135.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.*, *104*, 801–838.

Dick, F., Bates, E., Wulfeck, B., Utman, J., Dronkers, N., & Gernsbacher, M. (2001). Language deficits, localiztion, and grammar: Evidence for a distributive model of langauge breakdown in aphasic patients and neurologically intact individuals. *Psychol. Rev.*, *108*, 759–788.

Drai, D., & Grodzinsky, Y. (1999). Comprehension regularity in Broca's aphasia? There's more of it than you ever imagined. *Brain Lang.*, *70*, 139–143.

Dronkers, N., Wilkin, D., Van Valin, R., Redfern, B., & Jaeger, J. (2004). Lesion analysis of the brain areas involved in language comprehension *Cognition*, *92*, 145–177.

Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., Jr., & McFarlane, D. K. (1996). Effects of lexical frequency and syntactic complexity in spoken language comprehension: Evidence from the auditory moving window technique. *J. Exp. Psychol. Learn. Mem. Cogn.*, *22*, 324–335.

Fiebach, C. J., Schlesewsky, M., & Lohmann, G. (2005). Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Hum. Brain Mapp.*, *24*, 79–91.

Fiebach, J., Vos, S. H., & Friederici, A. D. (2004). Neural correlates of syntactic ambiguity in sentence comprehension for low and high span readers. *J. Cogn. Neurosci.*, *16*, 1562–1575.

Fodor, J. A. (1972). *The modularity of mind.* Cambridge, MA: MIT Press.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *J. Psycholinguist. Res.*, *18*, 51–60.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure.* Chicago: Chicago University Press.

Goldberg, A. (2006). *Constructions at work.* Oxford, UK: Oxford University Press.

Grewe, T. G., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D. Y., & Schlesewsky, M. (2005). The emergence of the unmarked: A new perspective on the language-specific function of Broca's area. *Hum. Brain Mapp.*, *26*, 178–190.

Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D. Y., & Schlesewsky, M. (2006). Linguistic prominence and Broca's area: The influence of animacy as a linearization principle. *NeuroImage*, *32*, 1395–1402.

Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D. Y., & Schlesewsky, M. (2007). The role of the posterior superior temporal sulcus in the processing of unmarked transitivity. *NeuroImage*, *35*, 343–352.

Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca'a area. *Behav. Brain Sci.*, *23*, 47–117.

Grodzinsky, Y., & Friederici, A. (2006). Neuroimaging of syntax and syntactic processing. *Curr. Opin. Neurobiol.*, *16*, 240–246.

Grodzinsky, Y., Wexler, K., Chien, Y.-C., Marakovitz, S., & Solomon, J. (1993). The breakdown of binding relations. *Brain Lang.*, *45*, 396–422.

Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1997). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain Lang.*, *59*, 76–120.

Haarmann, H. J., & Kolk, H. H. (1991). Syntactic priming in Broca's aphasics: Evidence for slow activation. *Aphasiology*, *5*, 247–263.

Haarmann, J., & Kolk, H. (1994). On-line sensitivity to subject-verb agreement violations in Broca's aphasics: The role of syntactic complexity and time. *Brain Lang.*, *46*, 493–516.

Hildebrandt, N., Caplan, D., & Evans, K. (1987). The man$_i$ left$_i$ without a trace: A case study of aphasic processing of empty categories. *Cogn. Neuropsychol.*, *4*(3), 257–302.

Indefrey, P., Hagoort, P., Herzog, H., Seitz, R., & Brown, C. (2001). Syntactic processing in the left prefrontal cortex in independent of lexical meaning. *NeuroImage*, *14*, 546–555.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, *274*, 114–116.

Karbe, H., Herholz, K., Szelies, B., Pawlik, G., Wienhard, K., & Heiss, W. (1989). Regional metabolic correlates of Token test results in cortical and subcortical left hemispheric infarction. *Neurology*, *39*, 1083–1088.

Kempler, D., Curtiss, S., Metter, E., Jackson, C., & Hanson, W. (1991). Grammatical comprehension, aphasic syndromes and neuroimaging. *J. Neurolinguistics*, *6*, 301–318.

King, J. W., & Just, M. A. (1991). Individual difference in syntactic processing: The role of working memory. *J. Mem. Lang.*, *30*, 580–602.

Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, *13*, 361–392.

MacDonald, M.-E., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.*, *101*, 676–703.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman.

Martin, R. C., & He, T. (2004). Semantic short-term memory deficit and language. *Brain Lang.*, *89*, 76–82.

Mesulam, M.-M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann. Neurol.*, *28*(5), 597–613.

Mesulam, M.-M. (1998). From sensation to cognition. *Brain*, *121*, 1013–1052.

Miyake, A. K., Carpenter, P., & Just, M. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like brain-damaged patients. *Cogn. Neuropsychol.*, *11*, 671–717.

Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H., Duncan, G. W., & Davis, K. R. (1978). Broca aphasia: Pathologic and clinical. *Neurology*, *28*, 311–324.

Page, M. P. (2006). What can functional imaging tell the experimental psychologist? *Cortex*, *42*, 428–443.

Pearlmutter, N., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *J. Mem. Lang.*, *34*, 521–542.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain Lang.*, *52*, 452–473.

Swinney, D., & Zurif, E. (1995). Syntactic processing in aphasia. *Brain Lang.*, *50*, 225–239.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Tramo, M. J., Baynes, K., & Volpe, B. T. (1988). Impaired syntactic comprehension and production in Broca's aphasia: CT lesion localization and recovery patterns. *Neurology*, *38*, 95–98.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *J. Mem. Lang.*, *33*, 285–318.

Tyler, L. (1985). Real-time comprehension processes in agrammatism: A case study. *Brain Lang.*, *26*, 259–275.

Vanier, M., & Caplan, D. (1989). CT-Scan correlates of agrammatism. In L. Menn & L. K. Obler (Eds.), *Agrammatic aphasia* (pp. 37–114). New York: J. Benjamin.

Waters, G. S., Caplan, D., Stanzcak, L., & Alpert, N. (2003). Individual differences in rCBF correlates of syntactic processing in sentence comprehension: Effects of working memory and speed of processing. *NeuroImage*, *19*, 101–112.

Zurif, E., Swinney, D., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. *Brain Lang.*, *45*, 448–464.

56 Semantic Unification

PETER HAGOORT, GIOSUÈ BAGGIO, AND ROEL M. WILLEMS

ABSTRACT Language and communication are about the exchange of meaning. A key feature of understanding and producing language is the construction of complex meaning from more elementary semantic building blocks. The functional characteristics of this semantic unification process are revealed by studies using event-related brain potentials. These studies have found that word meaning is assembled into compound meaning in not more than 500 ms. World knowledge, information about the speaker, co-occurring visual input, and discourse all have an immediate impact on semantic unification and trigger electrophysiological responses that are similar to those triggered by sentence-internal semantic information. Neuroimaging studies show that a network of brain areas, including the left inferior frontal gyrus, the left superior/middle temporal cortex, the left inferior parietal cortex, and, to a lesser extent, their right-hemisphere homologues are recruited to perform semantic unification.

Ultimately, language is the vehicle for the exchange of meaning between speaker and listener, between writer and reader. The unique feature of this vehicle is that it enables the assembly of complex expressions from simpler ones. The cognitive architecture necessary to realize this expressive power is tripartite in nature, with levels of form (sound, graphemes, manual gestures in sign language), syntax, and meaning as the core components of our language faculty (Jackendoff, 1999, 2002; Levelt, 1999). The principle of compositionality is often invoked to characterize the expressive power of language at the level of meaning. The most strict account of compositionality states that the meaning of an expression is a function of the meanings of its parts and the way they are syntactically combined (Fodor & Lepore, 2002; Heim & Kratzer, 1998; Partee, 1984). In this account, complex meanings are assembled bottom-up from the meanings of the lexical building blocks by means of the combinatorial machinery of syntax. This process is sometimes referred to as simple composition (Jackendoff, 1997). That this is not without problems can be seen in adjective-noun constructions such as "flat tire," "flat beer," "flat note," and so on (Keenan, 1979). In all these cases, the meaning of "flat" is quite different and strongly context dependent. For this and other reasons, simple composition seems not to hold across all possible expressions in the language (for a discussion of this and other issues related to compositionality, see Baggio, van Lambalgen, & Hagoort, in press). One of the challenges for a cognitive neuroscience of language is to account for the functional and neuroanatomical underpinnings of online meaning composition.

In linking the requirements of the language system as instantiated in the finite and real-time machinery of the human brain to the broader domain of cognitive neuroscience, three functional components are considered to be the core of language processing (Hagoort, 2005). The first is the *memory* component, which refers to the different types of language information stored in long-term memory (the mental lexicon) and to how this information is retrieved (lexical access). The *unification* component refers to the integration of lexically retrieved information into a representation of multiword utterances, as well as the integration of meaning extracted from nonlinguistic modalities; this component is at the heart of the combinatorial nature of language. Finally, the *control* component relates language to action, and is invoked, for instance, when the correct target language has to be selected (in the case of bilingualism) or for handling turn taking during conversation. In principle, this MUC (memory, unification, control) framework applies to both language production and language comprehension, although details of their functional anatomy within each component will be different. The focus of this chapter is on the unification component.

Classically, psycholinguistic studies of unification have focused on syntactic analysis. However, as we saw, unification operations take place not only at the syntactic processing level. Combinatoriality is a hallmark of language across representational domains (cf. Jackendoff, 2002). Thus, also at the semantic and phonological levels, lexical elements are combined and integrated into larger structures (cf. Hagoort, 2005). In the remainder of this chapter, we will discuss semantic unification. Semantic unification refers to the integration of word meaning into an unfolding representation of the preceding context. This is more than the concatenation of individual word meanings, as is clear from the adjective-noun examples given earlier. In the interaction with the preceding sentence or discourse context, the appropriate meaning is selected or constructed, so that a coherent interpretation results.

PETER HAGOORT Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen; Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
GIOSUÈ BAGGIO and ROEL M. WILLEMS Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

Hereafter we will first discuss the functional characteristics of semantic unification as revealed by ERP and MEG studies. Next, results from fMRI studies will be discussed to identify the neural networks of semantic unification. In the remainder we will use the terms *unification* and *integration* interchangeably. However, in the last paragraph we propose to use the terms *integration* and *unification* for two different ways of combining information.

Functional characteristics of semantic unification

Insights into the functional characteristics of semantic unification have been especially gained through a series of event-related potential (ERP) studies. Most studies on semantic unification exploit the characteristics of the so-called N400 component in the ERP waveform. Kutas and Hillyard (1980) were the first to observe this negative-going potential with an onset at about 250 ms and a peak around 400 ms (hence the N400), whose amplitude was increased when the semantics of the eliciting word (i.e., *socks*) mismatched with the semantics of the sentence context, as in "He spread his warm bread with socks."

Since its original discovery in 1980, much has been learned about the processing nature of the N400 (for extensive overviews, see Kutas & Federmeier, 2000; Kutas, Van Petten, & Kluender, 2006; Osterhout, Kim, & Kuperberg, 2007). In particular, as Kutas and Hillyard (1984) and many others have observed, the N400 effect does not depend on a semantic violation. For example, subtle differences in semantic expectancy, as between *mouth* and *pocket* in the sentence context "Jenny put the sweet in her mouth/pocket after the lesson," can also modulate the N400 amplitude (Hagoort & Brown, 1994). Specifically, as the degree of semantic fit between a word and its context increases, the amplitude of the N400 decreases. This general relation between individual word meanings and the semantics of the context is independent of type of context. That is, it is found for a single-word context (Holcomb, 1993), for a sentence context (Kutas & Hillyared, 1980, 1984), and for larger discourses (van Berkum, Hagoort, & Brown, 1999). Because of such subtle modulations, the N400 is generally taken to reflect processes involved in the integration of the meaning of a word into the overall semantic representation constructed for the preceding language input (Brown & Hagoort, 1993; Osterhout & Holcomb, 1992). However, different views exist as to what brings about the N400 integration effect. Federmeier and Kutas (1999; Kutas & Federmeier, 2000) proposed that in addition to its sensitivity to context, the N400 is also sensitive to the ease of accessing information from semantic memory. As such, the N400 can be seen to reflect the organization of (lexical) meaning in semantic memory. According to this view, the N400 amplitude is modulated by the degree to which the context contains retrieval cues for accessing or selecting the stored representation for a particular word meaning. Recent evidence in favor of this position was obtained in a study by DeLong, Urbach, and Kutas (2005). These authors found an N400 effect to an indefinite article (*an* versus *a*) that excluded the semantically expected continuation, such as in "the day was breezy so the boy went out to fly an . . . ," where *kite* would be the contextually expected noun. This result suggests a contextual preactivation of the target word. However, other recent evidence is more compatible with a unification account. Li, Hagoort, and Yang (2008) investigated the neurophysiological response to manipulations of information structure. An important distinction at the level of semantic/conceptual structure is that between conceptual content and information structure. The latter refers to the division of the content of a sentence into information that is in the foreground or in the background (topic/focus; given/new). In many languages new information is accented, whereas old information is deaccented. Li and colleagues found that in Chinese the N400 to new, accented information was larger than the N400 to new, deaccented information, despite the fact that the accentuation was contextually appropriate, whereas the absence of an accent was not. The authors argue that this result is best explained by the recruitment of additional unification resources for information that is marked as more salient by accentuation.

One way to reconcile these different accounts of the N400 is by reference to different roles for the left and right hemispheres (Kutas & Federmeier, 2000; Federmeier, 2007). Federmeier and Kutas (1999) did a visual-half-field study in which participants read sentences such as "Every morning John makes himself a glass of freshly squeezed juice. He keeps his refrigerator stocked with (oranges/apples/carrots)." In this context, "oranges" is the expected continuation, "apples" is a violation but within the correct semantic category, and "carrots" is a violation that crosses the category boundary. The left-visual-field/right-hemisphere (LVF/RH) results showed a smaller N400 to oranges than to both within- and across-category violations, but no N400 difference for the two types of violation. In contrast, for the RVF/LH a reduced N400 was obtained not only for the predicted word ("orange"), but also in part for the within-category violation ("apple") (see figure 56.1). This latter result can be explained as a consequence of a contextual prediction for the target concept. Owing to the organization of semantic memory, the within-category nontarget ("apple") gets activated to some degree as well, resulting in a partially reduced N400. Predictive semantic processing might thus be a left-hemisphere processing mechanism, while the right-hemisphere contribution is presumably strictly postlexical in nature, only contributing to the integration of the word meaning from a lexical item that received bottom-up support on the basis of visual or acoustic input.

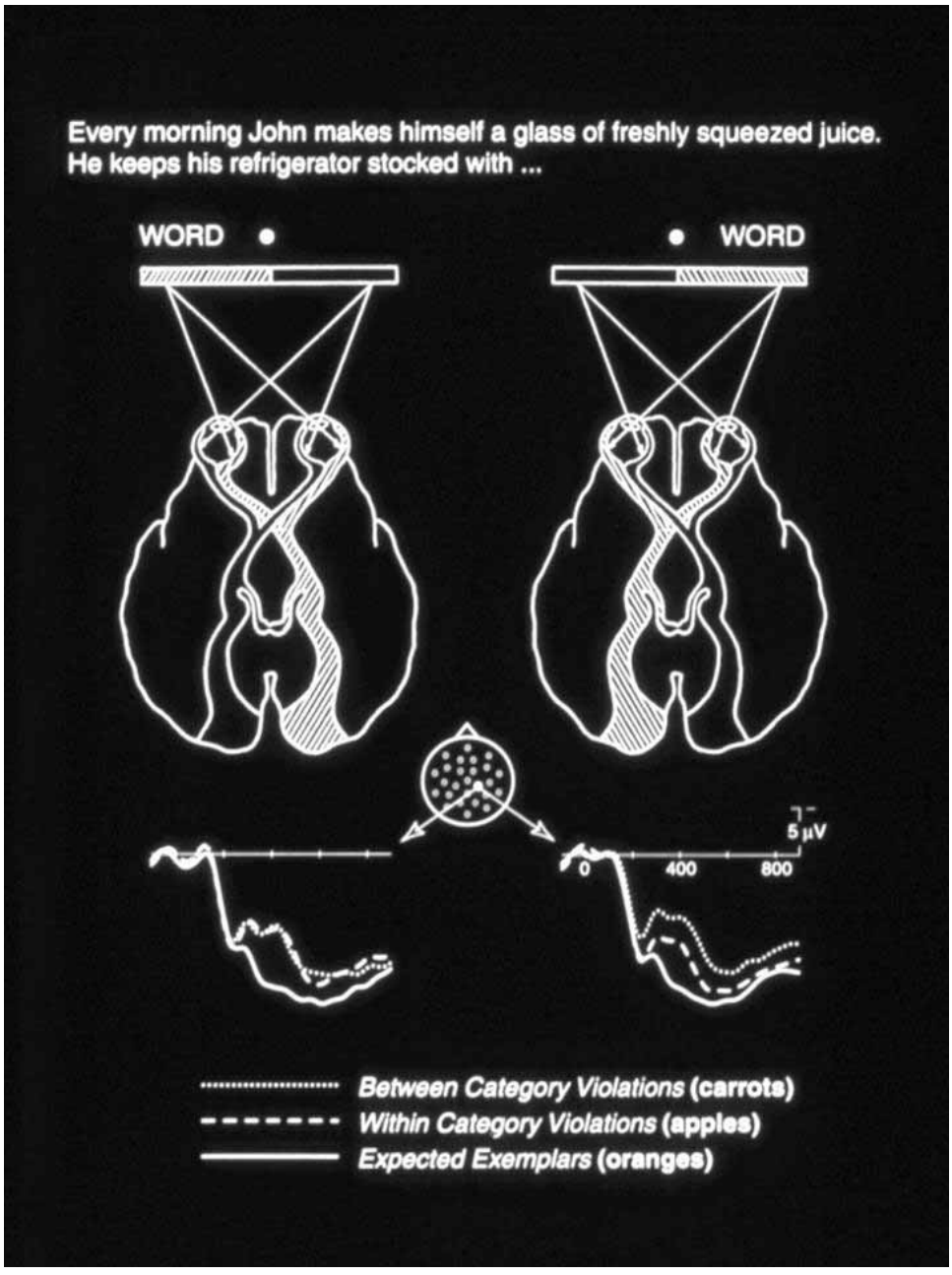


FIGURE 56.1 Participants read the sentences as in the example in a visual-half-field presentation design. Context words were presented at central fixation, whereas sentence-final target words (e.g., "oranges") were presented to the left or right of fixation. As illustrated, words presented to the left visual field (LVF) travel initially to the right hemisphere (RH) and vice versa. ERPs are shown here from a representative (right medial central) site as indicated. The response to target words presented to the RVF (left hemisphere) (shown on right), yielded the same pattern as that observed with central fixation: expected exemplars (solid line) elicited smaller N400s than did violations of either type, but within-category violations (dashed line) also elicited smaller N400s than between-category violations (dotted line). This pattern is indicative of a "predictive" strategy, in which semantic information associated with the expected item is preactivated in the course of processing the context information. The response to targets presented to the LVF/RH (shown on left), however, was qualitatively different: expected exemplars again elicited smaller N400s than violations, but the response to the two types of violations did not differ. This pattern is more consistent with a plausibility-based integrative strategy. Taken together, the results indicate that the hemispheres differ in how they use context to process semantic information in online language processing. (Reprinted with permission from Kutas & Federmeier, 2000.)

In recent years, the N400 and other language-relevant ERP effects have been exploited to test more specific ideas about the functional characteristics of semantic unification. These include the contribution of world knowledge, the processing of silent meaning, the integration of pragmatic information, and the syntax-semantics interface. We will discuss briefly each of these theory-driven issues.

World Knowledge At least since Frege (1892; see Seuren, 1998), theories of meaning make a distinction between the semantics of an expression and its truth-value in relation to our mental representation of the state of affairs in the world (Jackendoff, 2002). For instance, the sentence "Bill Clinton is the 43rd president of the USA" has a coherent semantic interpretation, but contains a proposition that is false in the light of our knowledge that George W. Bush is the 43rd president. The situation is different for the sentence "The presidential helicopter is divorced." Under default interpretation conditions, this sentence has no coherent semantic interpretation, since the predicate "is divorced" requires an animate argument. The difference between these two sentences points to the distinction that can be made between facts of the world ("world knowledge") and facts of the words of our language, including their meaning ("linguistic knowledge"). Hagoort, Hald, Bastiaansen, and Petersson (2004) performed a combined EEG/fMRI study that compared the unification of linguistic knowledge with the unification of world knowledge. While participants' brain activity was recorded, they read one of three versions of a sentence such as "The Dutch trains are *yellow*/*white*/*sour* and very crowded" (critical words are in italics). It is a well-known fact among Dutch people that Dutch trains are yellow, and therefore the first version of this sentence is correctly understood as true. However, the linguistic meaning of the alternative color term *white* applies equally well to trains as the predicate *yellow*. It is world knowledge about trains in Holland that makes the second version of this sentence false. This is different for the third version, where (under standard interpretation conditions) the core semantic features of the predicate *sour* do not fit the semantic features of its argument *trains*.

Figure 56.2 presents an overview of the results. As expected, the classic N400 effect was obtained for the semantic violations. For the world-knowledge violations, a clear N400 effect was observed as well. Crucially, this effect was identical in onset and peak latency, and very similar in amplitude and topographic distribution to the semantic N400 effect. This finding is strong empirical evidence that lexical-semantic knowledge and general world knowledge are both integrated in the same time frame during sentence interpretation. The results of this world-knowledge experiment provide further evidence against an account of unification in which first the meaning of a sentence is determined, and only then is its meaning verified in relation to our knowledge of the world. Semantic interpretation is not separate from its integration with nonlinguistic conceptual knowledge.

Further evidence in favor of an enriched composition account comes from a study on the integration of information about the speaker. In interpreting a speaker's utterance, we take not only the preceding utterances into consideration, but also our knowledge of the speaker. For instance, we might find it odd for a man, but not for a woman of a certain age, to say, "I think I am pregnant." At some point during language comprehension, the listener combines the information that is represented in the content of a sentence with the information she has about the speaker. The question is: When exactly does the pragmatic information about the speaker have its impact on the unfolding interpretation of the utterance? This question was answered in a recent ERP study by van Berkum, van den Brink, Tesink, Kos, and Hagoort (2008). Participants listened to sentences, some of which contained a specific word at which the message content became at odds with inferences about the speaker's sex, age, and social status, as inferred from the speaker's voice.

If voice-based inferences about the speaker are recruited by the same early unification process that combines word meanings, then speaker inconsistencies and semantic anomalies should elicit the same N400 effect. This was indeed observed. Reliable effects of speaker inconsistency were already found in the 200–300-ms latency range after word onset. The same latency effect was obtained for the straightforward semantic anomalies. These findings therefore demonstrate that sense making depends on the pragmatics of the communicative situation right from the start.

As for compositionality, the results of the studies just reviewed may mean two things, depending on one's views on the lexicon. One possibility is that the lexicon includes declarative memory in its entirety, and then simple composition seems enough to account for the similarity between the N400 effects. Alternatively, the lexicon includes invariant (i.e., linguistic) meanings only, and then enriched composition—the thesis that the lexicon is not the only source of semantic content—seems necessary to explain the observed N400 effects (Baggio et al., in press).

Event Knowledge and Discourse Models Unification of lexical representations ultimately results in a discourse model—that is, a representation making what is given as input true whenever possible (recall the Dutch trains examples). Events offer a vantage point for investigating the properties of discourse models, because natural languages have very sophisticated devices for characterizing time and causation. One of these devices is aspect. This is the linguistic marking of the internal profile of events. Ferretti, Kutas, and McRae (2007) found that readers have least difficulty

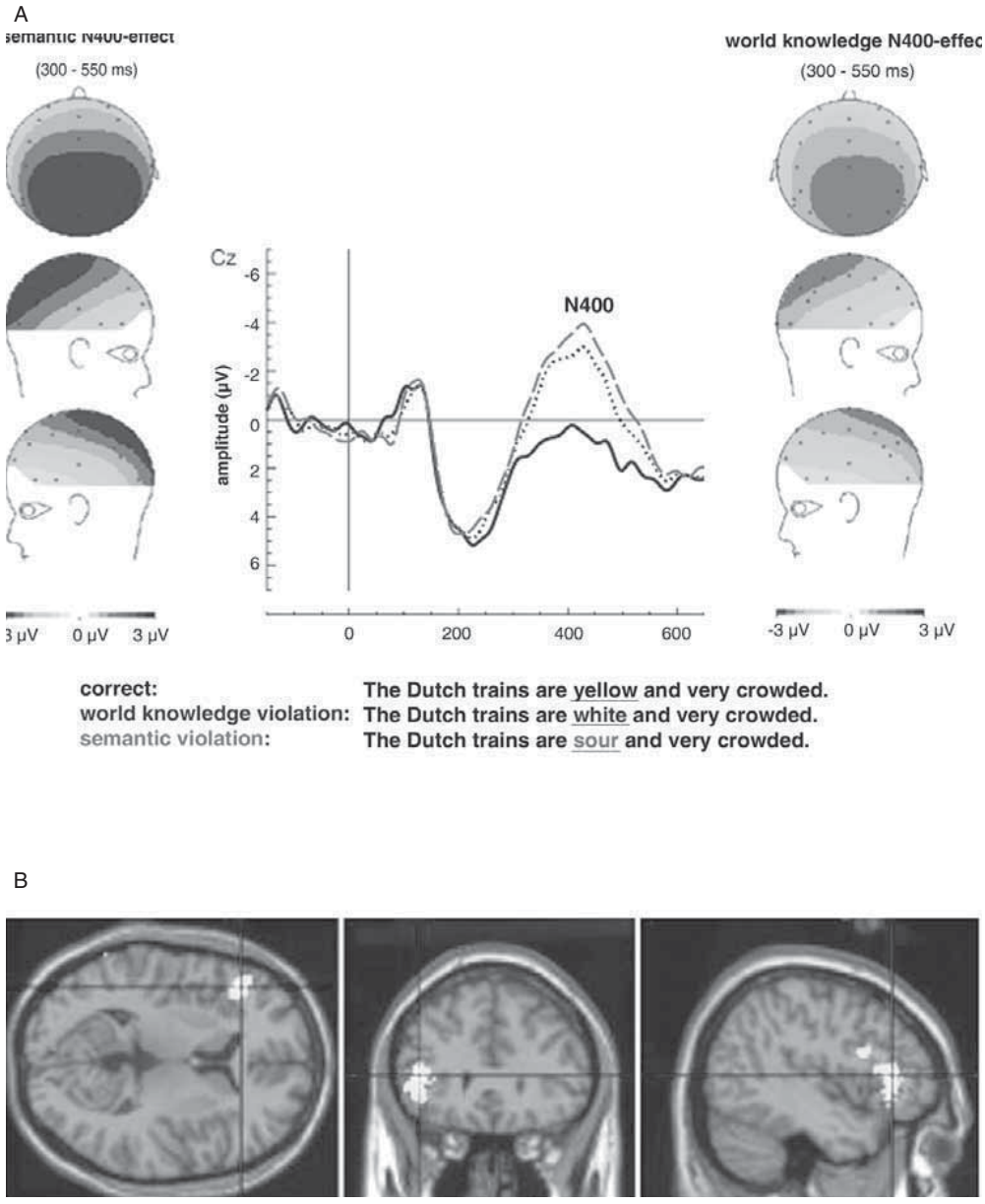


FIGURE 56.2 (*A*) Grand average ERPs for a representative electrode site (Cz) for correct condition (black line), world-knowledge violation (blue dotted line), and semantic violation (red dashed line). ERPs are time locked to the presentation of the critical words (underlined). Spline-interpolated isovoltage maps display the topographic distributions of the mean differences from 300 to 550 ms between semantic violation and control (left), and between world knowledge violation and control (right). (*B*) The common activation for semantic and world-knowledge violations compared to the correct condition, based on the results of a minimum-T-field conjunction analysis. Both violations resulted in a single common activation ($P = 0.043$, corrected) in the left inferior frontal gyrus. The crosshairs indicate the voxel of maximal activation. (Reprinted with permission from Hagoort, Hald, Bastiaansen, & Petersson, 2004.) (See color plate 71.)

integrating locative nouns when the aspect of the main verb is imperfective and the denoted location is a prototypical one given the verb's semantics. In sentences with an imperfective, such as "The diver was snorkeling in the ocean/pond," a larger N400 was evoked by *pond* than by *ocean*. This N400 effect was reduced if the aspect was perfective, as in "The diver had snorkeled in the ocean/pond." Describing an event as ongoing using the imperfective aspect leads readers to construct a situation model in which locations and other dimensions of the action become relevant, while such dimensions are ignored if the action is viewed perfectively.

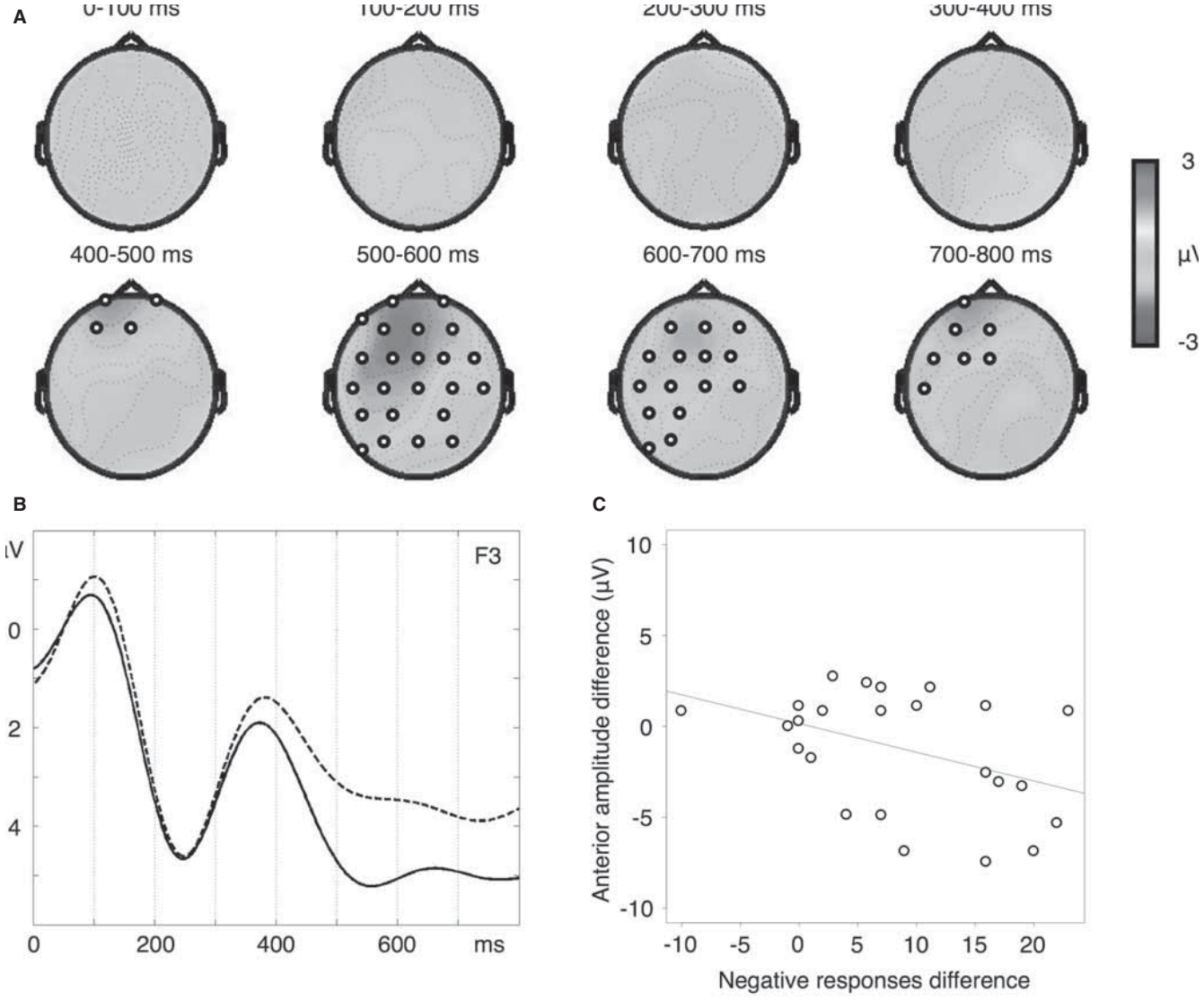
The imperfective leads also to expectations concerning the outcome of the event described. Baggio, van Lambalgen, and Hagoort (2008) investigated whether, in sentences like "The girl was writing a letter when her friend spilled coffee on the tablecloth/paper," the goal state (a complete letter)

was represented online during the unification process. If the goal is predicted to occur whenever the imperfective is used, a difference should be observed at the word *paper* compared to *tablecloth*. Spilling coffee on the paper implies that the goal state was not attained, and forces the system to revise the earlier commitment to the event's completion (Baggio & van Lambalgen, 2007). Spilling coffee on the tablecloth, however, does not have this implication. *Paper* did indeed result in a larger sustained anterior negativity (SAN) compared to *tablecloth*, and the effect was correlated with the frequency with which participants concluded that the event was not completed (see figure 56.3). These results again suggest that semantic processing is not bound to asserted content, but can include inferences anticipating the outcome of actions and events, as well as other inferences invalidating previously drawn conclusions. In this sense, unification can be described as a defeasible process: discourse models built up incrementally at any one stage may have to be revised when additional information becomes available, as when the word *paper* is encountered in this example (cf. Carreiras, Garnham, Oakhill, & Cain, 1996; Sturt, 2007).

Fictional Discourse and Silent Meaning Simple composition implies that unification preserves the semantic identity of the constituent expressions. However, experimental research suggests that discourse may override even such core features of word semantics as animacy. Nieuwland and van Berkum (2006) showed that sentences that make sense on their own, like "The peanut was salted," appear anomalous if they are embedded in a context in which the inanimate subject (the peanut) is attributed animate features. In a narrative in which the peanut danced and sang, because it fell in love with an almond it had met, the final word in "The peanut was salted" resulted in a larger N400 compared to "The peanut was in love" (see figure 56.4). This result is taken to show that discourse can override seemingly context-invariant semantic features of words.

Another interesting phenomenon is that of silent meaning—that is, meaning not expressed in the syntax and phonology of an expression. A number of linguistic devices are available to speakers and hearers that allow efficient communication of meaning beyond what is explicitly asserted. Among these are coercing expressions, functioning as a shorthand for lengthier definite descriptions, as in the classic examples "The ham sandwich in the corner wants some more coffee," where *ham sandwich* in fact refers to the person who ordered one, and "Plato is on the top shelf next to Russell," where *Plato* and *Russell* refer to copies of the works of the two philosophers. More extreme forms of coercion are possible, as in "Fishing the edges dry," where *dry* is a condensed expression for the phrase *using a dry fly*, or in resultative constructions like "Hammering the metal flat," where *flat* denotes the final state of the metal after hammering. What all these widely used expression types have in common is a silent semantic element, which has to be recovered (sometimes obligatorily) to make full sense of the sentence. Semantic processing might be taxed during such recovery process, and that is indeed what was found experimentally. Complement-coercing sentences like "The journalist began the article," which presumably means that she began writing or typing the article, are more difficult to process than sentences in which the activity is part of the asserted content like "The journalist wrote the article." The processing costs of complement coercion have been established using reading times (McElree, Traxler, Pickering, Seely, & Jackendoff, 2001), eye tracking (Traxler, Pickering, & McElree, 2002; Traxler, McElree, Williams, & Pickering, 2005), and MEG (Pylkkänen & McElree, 2007). Pylkkänen and McElree found a magnetoencephalography (MEG) response that was located in ventromedial prefrontal cortex to coerced sentences, which was different from the M350, the magnetic correlate of the N400. Semantic processing beyond the single-word level is therefore not restricted to processing asserted content as delivered by the input, but is crucially engaged in recovering silent meaning in presuppositions, implicatures, coercions, and so on. Crucially, recovered meanings are triggered by expressions that are given as input but are themselves phonologically and syntactically silent, an effect that shows that semantics is relatively independent from the two other components of the language system. This "autonomy of semantics" is at odds with the syntax-semantics homomorphism postulated by formal semanticists (Montague, 1970; Partee, Ter Meulen, & Wall, 1990), as well as with the "interface uniformity" upon which generative grammar is built (Culicover & Jackendoff, 2005).

Unification and the Syntax-Semantics Interface A language-relevant ERP effect that has been related to syntactic processing is a positivity, nowadays referred to as P600 or as P600/SPS (Coulson, King, & Kutas, 1998; Hagoort, Brown, & Osterhout, 1999; Osterhout, McLaughlin, & Bersick, 1997). The P600 is the syntactic equivalent of the N400 effect. One of the antecedent conditions of P600 effects is a violation of a syntactic constraint. The relation between N400 and the P600 effects might provide insights into the interplay between semantic and syntactic unification. Modulations of the P600 have been observed not only to syntactic violations, syntactic ambiguities, and syntactic complexity, but also to breakdowns of normal operations at the syntax-semantics interface (for a review, see Kuperberg, 2007). For example, Kim and Osterhout (2005) reported larger P600s evoked by *devouring* in "The hearty meal was devouring . . . ," compared to either "The hearty meal was devoured . . ." or "The hungry boys were devouring . . ."; this despite the fact that the sentence is syntactically well



Het meisje was een brief aan het schrijven toen haar vriendin koffie op het tafelkleed morste.

'The girl was writing a letter when her friend spilled coffee on the tablecloth.'

Het meisje was een brief aan het schrijven toen haar vriendin koffie op het papier morste.

'The girl was writing a letter when her friend spilled coffee on the paper.'

FIGURE 56.3 (*A*) Grand-average topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final verb when it terminated versus when it did not terminate the accomplishments in the progressive. Circles represent electrodes in a significant ($P < 0.05$) cluster. (*B*) Grand-average ERP waveforms from a representative site (F3) time-locked to the onset (0 ms) of the verb in terminated versus nonterminated accomplishments. Negative values are plotted upward. (*C*) Scatter plot displaying the correlation between the amplitude of the sustained anterior negativity elicited by terminated accomplishments and the frequency of negative responses in a button-press, probe-selection task ($r = -0.415$, $T(22) = -2.140$, $P = 0.043$). The mean difference of negative responses between terminated and nonterminated accomplishments is plotted on the abscissa. The mean amplitude difference at frontopolar and frontal electrodes between terminated and nonterminated accomplishments in the 500–700-ms interval following the onset of the sentence-final verb is plotted on the ordinate. (After Baggio, van Lambalgen, & Hagoort, 2008.) (See color plate 72.)

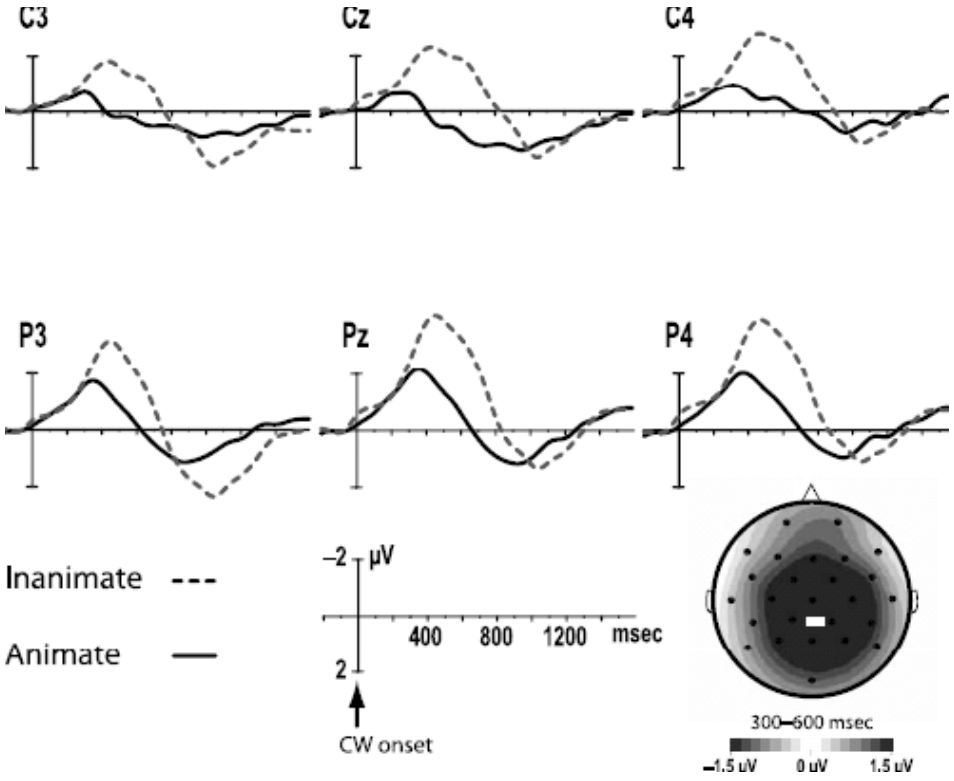


FIGURE 56.4 N400 effects triggered by a correct predicate (*salted*) that is, however, contextually disfavored in comparison to an incorrect predicate (*in love*). Waveforms are presented for representative electrode sites, time-locked to the onset of the critical inanimate/animate predicate in the fifth sentence. (After Nieuwland & van Berkum, 2006.)

formed (see figure 56.5). The semantics of *meal* and *devour* suggest a plausible thematic role assignment to *meal*: a *theme* instead of an *agent* as the syntax implies. In this case, semantic plausibility overrides syntactic constraints, and the verb *devouring* is presumably perceived as a morphosyntactic violation indexed by the P600. Conflicts between syntactic and semantic constraints might result in N400 or P600 effects depending on whether, respectively, the semantic or the syntactic constraints are the weakest. In cases where the input is anomalous because of a conflict between semantic and syntactic cues, the modus operandi of the system seems to obey a "loser takes all" principle. That is, if the semantic cues are stronger than the syntactic cues, the effect will appear at the level of syntactic unification (P600). Kuperberg (2007) argues that there are at least two neural routes subserving language comprehension: (1) a semantic, memory-based stream that provides elementary meanings as well as conceptual, categorical, and thematic relations between them; (2) a combinatorial stream that provides analyses based on morphosyntactic constraints and thematic roles as given in the input. The P600 reported by Kim and Osterhout (2005), for example, might be taken to suggest that semantic associations between words are the strongest constraints—for instance, because in this case they are taken into account earlier than the syntactic cues.

CONCLUSION In general, ERP research on semantic processing has found that word meaning is very rapidly assembled into compound meaning. This statement holds for individual word meanings in the context of single words, sentences, or discourse. But it also holds for meaning that is extracted from pictures, co-speech gestures, or stereotypes inferred from speaker characteristics (Willems, Özyürek, & Hagoort, 2007, 2008; van Berkum et al., 2008). The effects of semantic processing are most often observed as modulations of the N400 amplitude. The topographic distribution of the N400 differs slightly for different stimulus types. It is more evenly distributed for auditory than for the visual N400. Pictures and co-speech gestures elicit a more frontal N400 than sentences without concomitant nonlinguistic information. This finding suggests that the set of neural generators contributing to the scalp-recorded N400 is not fully overlapping for the different types of meaningful stimuli. This result is consistent with the results from fMRI studies, showing both overlapping and distinct activations in connection to the various types of meaningful input (see the next section). Intracranial recordings and MEG studies indicate that the scalp-recorded N400 is caused by coordinated activity in a number of different brain areas, including the anterior inferotemporal cortex (McCarthy, Nobre, Bentin, & Spencer, 1995), the superior temporal cortex

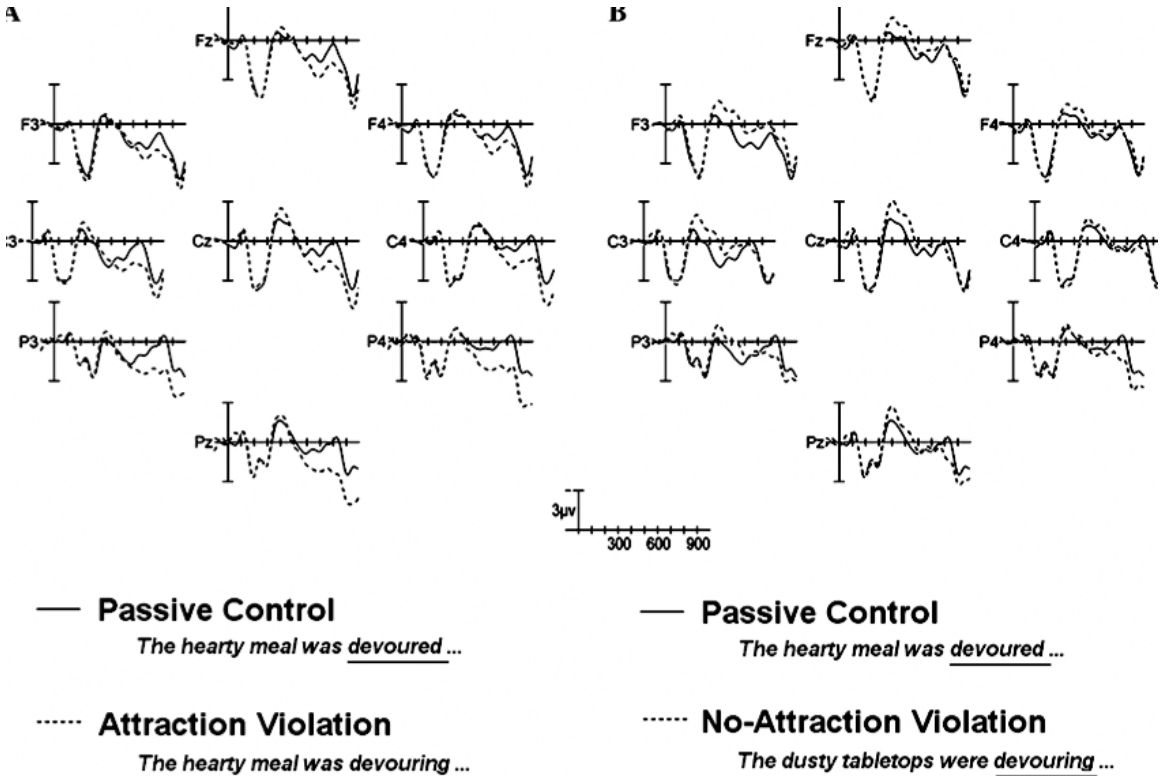


FIGURE 56.5 At the interface between syntax and semantics. Grand-average ERPs recorded at three midline sites and six medial-lateral sites. All sentences are syntactically correct. (*A*) ERPs to passive control verbs (solid line) and thematic violation verbs (dashed line). (*B*) ERPs to active control verbs (solid line) and thematic violation verbs (dashed line). In both cases the inconsistency between grammatical roles and thematic role biases resulted in robust P600 effects. Onset of the critical verbs is indicated by the vertical bar. Each hash mark represents 100 ms. Positive voltage is plotted down. (Kim & Osterhout, 2005; reprinted with permission.)

(Dale et al., 2000; Helenius, Salmelin, Service, & Connolly, 1998; Halgren et al., 2002), and the left inferior frontal cortex (Halgren et al., 1994, 2002; Guillem, Rougier, & Claverie, 1999). Other ERP effects (e.g., anterior negativities) have also been observed to aspects of postlexical semantic processing. How they differ from the N400 effects in their functional characterization is an issue for further research.

The semantic unification network

In recent years a series of fMRI studies were aimed at identifying the semantic unification network. These studies either compared sentences containing semantic/pragmatic anomalies with their correct counterparts (Hagoort et al., 2004; Newman, Pancheva, Ozawa, Neville, & Ullman, 2001; Kuperberg et al., 2000, 2003; Kuperberg, Sitnikova, & Lakshmanan, 2008; Ni et al., 2000; Baumgaertner, Weiller, & Buchel, 2002; Kiehl, Laurens, & Liddle, 2002; Friederici, Ruschemeyer, Hahne, & Fiebach, 2003; Ruschemeyer, Zysset, & Friederici, 2006) or compared sentences with and without semantic ambiguities (Hoenig & Scheef, 2005; Rodd, Davis, & Johnsrude, 2005; Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007; Davis et al., 2007). The most consistent finding across all these studies is the activation of the left inferior frontal cortex (LIFC), more particularly BA 47 and BA 45. In addition, the left superior and middle temporal cortex is often found to be activated (see figure 56.6 for an overview), as well as left inferior parietal cortex. For instance, Rodd and colleagues had subjects listen to English sentences such as "There were dates and pears in the fruit bowl" and compared to the BOLD response of these sentences to the BOLD response of sentences such as "There was beer and cider on the kitchen shelf." The crucial difference between these sentences is that the former contains two homophones—"dates" and "pears"—which, when presented auditorily, have more than one meaning. This is not the case for the words in the second sentence. The sentences with the lexical ambiguities led to increased activations in LIFC and in the left posterior middle/inferior temporal gyrus. In this experiment all materials were well-formed English sentences in which the ambiguity usually goes unnoticed. Nevertheless, the results were very similar to those obtained in experiments that used semantic anomalies. Areas involved in semantic unification were found to be sensitive to the increase in semantic unification load that resulted from the ambiguous words.

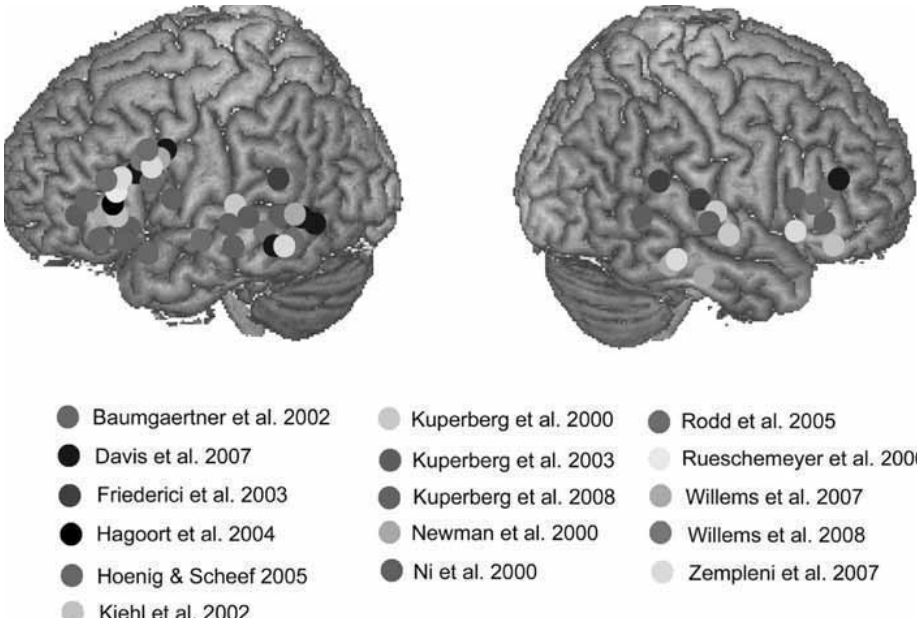


FIGURE 56.6 Overview of local maxima in inferior frontal cortex and in temporal cortex in neuroimaging studies employing sentences with semantic anomalies or semantic ambiguities. The local maxima (in MNI space) of each study were overlaid on a rendering of a brain in MNI space. For local maxima see tables 56.1 and 56.2; for a summary of the results see table 56.3. Rendering was made using MRIcroN. Please note that the local maxima of the Ni and colleagues (2000) and the Kuperberg and colleagues (2003) studies are displayed, but that these are not based on coordinates, since no coordinates were provided. The local maxima are drawn by hand based upon the figures in the respective papers. (See color plate 73.)

In short, the semantic unification network seems to include at least LIFC, left superior/middle temporal cortex, and the (left) inferior parietal cortex. To some degree, the right hemisphere homologues of these areas are also found to be activated (see figure 56.6). In the following subsections we will discuss the possible contributions of these regions to semantic unification.

THE MULTIMODAL NATURE OF SEMANTIC UNIFICATION An indication for the respective functional roles of the left frontal and temporal cortices in semantic unification comes from a few studies investigating semantic unification of multimodal information with language. Using fMRI, Willems and colleagues assessed the neural integration of semantic information from spoken words and from co-speech gestures into a preceding sentence context (Willems et al., 2007). Spoken sentences were presented in which a critical word was accompanied by a co-speech gesture. Either the word or the gesture could be semantically incongruous with respect to the previous sentence context. Both an incongruous word and an incongruous gesture led to increased activation in LIFC as compared to congruous words and gestures (see Willems et al., 2008, for a similar finding with pictures of objects). Interestingly, the activation of the left posterior STS was increased by an incongruous spoken word but not by an incongruous hand gesture. The latter resulted in a specific increase in dorsal premotor cortex (Willems et al., 2007). This finding suggests that activation increases in left posterior temporal cortex are triggered most strongly by processes involving the retrieval of lexical-semantic information. LIFC, however, is a key node in the semantic unification network, unifying semantic information from different modalities.

From these findings it seems that semantic unification is realized in a dynamic interplay between LIFC as a multimodal unification site on the one hand, and modality-specific areas on the other hand.

SEMANTIC UNIFICATION BEYOND THE SENTENCE LEVEL Recently a few studies have set out to investigate the neural networks involved in semantic processing at the level of multisentence utterances, such as short stories. Besides the network that is also activated to semantic unification at the sentence level, story comprehension involves activation of dorsomedial prefrontal cortex and, presumably, right inferior frontal cortex. In a recent meta-analysis, Ferstl and colleagues report the consistent involvement of medial prefrontal cortex, left STS/MTG, and LIFC when participants process coherent text as compared to sentences that do not form a coherent story or as compared to word lists (Ferstl, Neumann, Bogler, & von Cramon, 2008). In a variant of this line of research, Kuperberg, Lakshmanan, Caplan, and Holcomb (2006) presented participants with sentence quartets in which the relation of the last

TABLE 56.1

Involvement of the inferior frontal cortex in fMRI studies of sentence comprehension employing semantic anomalies or semantic ambiguities.

The table shows the studies that were used for the overview in figure 56.6, a brief description of the contrast that was employed in each of the studies, the reported coordinates of the local maxima in inferior frontal cortex in MNI space, and a verbal description of the location of the local maxima. When necessary, Talairach coordinates were converted to MNI space using the transformation suggested by Brett (http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach). Note that in computing the mean coordinates the findings from Kuperberg and colleagues (2003) and Ni and colleagues (2000) were not taken into consideration, since no coordinates were reported in these studies.

Study	Comparison	Coordinates *x y z* (MNI)	Region
Baumgaertner et al., 2002	Sem. incongruent > congruent	−51 36 −6	Left IFG
Davis et al., 2007	High ambiguity > low ambiguity	−40 24 18	Left IFG
		−48 6 34	
		−40 18 24	
		46 36 18	Right IFG
Friederici et al., 2003	Sem. incongruent > congruent	No activation	—
Hagoort et al., 2004	Sem. incongruent > congruent ∩ World knowledge incongruent > congruent	−44 30 8	Left IFG
Hoenig & Scheef, 2005	Sem. incongruent > congruent	−50 18 −14	Left IFG
		−50 43 11	
Kiehl et al., 2002	Sem. incongruent > congruent	−48 32 4	Left IFG/ant. temporal
		36 32 −16	Right IFG/ant. temporal
Kuperberg et al., 2000	Sem. incongruent > congruent	No activation	—
	Pragm. incongruent > congruent	No activation	—
Kuperberg et al., 2003	Pragm. incongruent > congruent	(No coordinates)	Left IFG
Kuperberg et al., 2008	Pragmatic incongruent > congruent	−43 25 −10	Left IFG
	Sem. incongruent > congruent	−49 4 10	Left IFG
		29 19 5	Right IFG
Newman et al., 2001	Sem. incongruent > congruent	−50 34 5	Left IFG
Ni et al., 2000	Sem. incongruence detection > tone pitch discrimination	(No coordinates)	Left IFG
			Right IFG
	Oddball paradigm with semantically incongruent sentences	(No coordinates)	Left IFG
		(No coordinates)	Right IFG
Rodd et al., 2005	High ambiguity > low ambiguity	−50 30 20	Left IFG
		−56 16 22	Left IFG
		−42 14 32	Left IFS
		36 26 4	Right IFG
		50 36 16	Right IFG
Rueschemeyer et al., 2006	Sem. incongruent > synt. incongruent	−50 30 15	Left ant. IFG
Willems et al., 2007	Sem. incongruent > congruent	−43 11 27	Left IFS
Willems et al., 2008	Sem. incongruent > congruent	−45 14 27	Left IFS
Zempleni et al., 2007	Subordinate meaning > dominant meaning	−48 26 20	Left IFG
		−52 16 26	Left IFG
		34 20 −10	Right IFG

TABLE 56.2

Involvement of the temporal cortex in fMRI studies of sentence comprehension employing semantic anomalies or semantic ambiguities.

The table shows the studies that were used for the overview in figure 56.6, a brief description of the contrast that was employed in each of the studies, the reported coordinates of the local maxima in temporal cortex in MNI space, and a verbal description of the location of the local maxima. When necessary, Talairach coordinates were converted to MNI space using the transformation suggested by Brett (http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach). Note that in computing the mean coordinates the findings from Kuperberg and colleagues (2003) and Ni and colleagues (2000) were not taken into consideration, since no coordinates were reported in these studies.

Study	Comparison	Coordinates *x y z* (MNI)	Region
Baumgaertner et al., 2002	Sem. incongruent > congruent	No activation	—
Davis et al., 2007	High ambiguity > low ambiguity	−50 −44 −12	Left ITG
		−54 −60 −2	
Friederici et al., 2003	Sem. incongruent > congruent	−60 −42 20	Left STG
		63 −40 20	Right STG
		58 −24 13	Right STG
Hagoort et al., 2004	Sem. incongruent > congruent	No activation	—
Hoenig & Scheef, 2005	Sem. incongruent > congruent	No activation	—
Kiehl et al., 2002	Sem. incongruent > congruent	No activation	—
Kuperberg et al., 2000	Sem. incongruent > congruent	43 −11 −7	Right MTG
		49 −17 4	Right STG
	Pragm. incongruent > congruent	−49 −31 9	Left STG
Kuperberg et al., 2003	Pragm. incongruent > congruent	(No coordinates)	Left STS
Kuperberg et al., 2008	Pragm. violations > correct sentences	−27 −28 −19	Left ant. med. temporal cortex
	Sem. incongruent > congruent	−53 −20 −1	Left STG
		58 −19 3	Right STG
Newman et al., 2001	Sem. incongruent > congruent	70 −36 −15	Right MTG
Ni et al., 2000	Sem. incongruence detection > tone pitch discrimination	(No coordinates)	Left STG/MTG
		(No coordinates)	Right STG/MTG
	Oddball paradigm with semantically incongruent sentences	(No coordinates)	Left pSTG
Rodd et al., 2005	High ambiguity > low ambiguity	−52 −50 −10	Left pITG
		−58 −8 −6	Left STG
Rueschemeyer et al., 2006	Sem. incongruent > synt. incongruent	—	—
Willems et al., 2007	Sem. incongruent > congruent	−53 −52 2	Left STS
Willems et al., 2008	Sem. incongruent > congruent	−53 −35 −3	Left STS
Zempleni et al., 2007	Subordinate meaning > dominant meaning	−50 −48 −12	Left ITG/MTG
		56 −34 −16	Right ITG/MTG

TABLE 56.3

Summary of the activations in the studies used for the overview in figure 56.6.

The coordinates from tables 56.1 and 56.2 were used. Table 56.3 specifies the mean coordinates for left and right inferior frontal and temporal cortices, the standard deviation in the *x*, *y*, and *z* directions in millimeters, the mean Euclidian distance of the local maxima to the mean coordinates, the number of maxima that were reported, and the number of studies that report maxima in that region. Note that the number of maxima is higher than the number of studies, since several studies report more than one maximum. Note that the findings from Kuperberg and colleagues (2003) and Ni and colleagues (2000) were not used in computing the mean coordinates, since no coordinates were reported in these studies.

	Mean (*x y z*) (MNI)	SD (*x y z*) (mm)	Mean Distance to Mean (mm)	Number of Studies (out of total)
Inferior frontal cortex				
Left	−47 22 14	4.3 10.6 13.9	16.3	14/16
Right	39 28 3	7.9 7.7 13.6	15.0	6/16
Temporal cortex				
Left	−51 −38 −3	8.6 15.4 10.9	18.0	10/16
Right	57 −26 0	8.8 10.9 13.7	17.2	6/16

sentence to the previous story context was manipulated. The less related sentences required an extra causal inference in order to make sense of the story. It was found that less related sentences (which evoked more inferencing) led to stronger activations in left and right IFC, left MTG, left middle fontal gyrus, and bilateral medial prefrontal cortex (Kuperberg et al.; see Hasson, Nusbaum, & Small, 2007, for a related result). These and other studies (e.g., St George, Kutas, Martinez, & Sereno, 1999; Xu, Kemeny, Park, Frattali, & Braun, 2005; Sieborger, Ferstl, & von Cramon, 2007) suggest that LIFC and left superior/middle temporal cortex are also important for unification of information beyond the sentence level. It is interesting to note that the medial prefrontal cortex, which is found activated for discourse but not for sentence-level processing, has been implicated in so-called mentalizing tasks, requiring the observer to take the perspective of someone else (Buckner, Andrews-Hanna, & Schacter, 2008; Frith & Frith, 2006). According to Mason and Just, this domain-general area is recruited in discourse processing for the sake of interpreting a protagonist's or agent's perspective (Mason & Just, 2006). In addition, right-hemisphere regions are sometimes but not consistently reported in the context of discourse processing (Maguire, Frith, & Morris, 1999; St George et al., 1999; Ferstl et al., 2008) (see Ferstl et al., 2008; Mason & Just, 2006, for extensive reviews). Some studies find that the temporal poles may be related to successful integration during story comprehension (Fletcher et al., 1995; Maguire et al.). The studies that report these activations are mostly done using PET. It is hard to assess the consistency of temporal pole activation during story/text comprehension because of the susceptibility to artifacts that these regions often suffer from in fMRI studies (but see Xu et al.; Ferstl et al.).

CONTROLLED PROCESSING AND SELECTION ACCOUNTS FOR LIFC Although LIFC (including Broca's area) has traditionally been construed as a language area, there is a wealth of recent neuroimaging data suggesting that its role extends beyond the language domain. Several authors have therefore argued that LIFC function is best characterized as "controlled retrieval" or "(semantic) selection" (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Wagner, Pare-Blagoev, Clark, & Poldrack, 2001; Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; Gold, Balota, Kirchoff, & Buckner, 2005; Moss et al., 2005; Thompson-Schill, Bedny, & Goldberg, 2005). For instance, Thompson-Schill and colleagues showed that LIFC was more strongly activated in a verb-generation task when the noun that served as the cue allowed for many different verb responses, as opposed to nouns that are reliably related to only one or a few verbs (Thompson-Schill et al., 1997). In response to the noun cue "scissors," for example, most participants generate the verb "to cut," whereas the noun "wheel" triggers a more diverse set of responses. On the basis of these and other findings, it was argued that LIFC guides semantic selection among competing alternatives, with higher activation when there are more competitors.

How does the selection account of LIFC function relate to the unification account? As is discussed in more detail elsewhere, unification often implies selection (Hagoort, 2005). For instance, in the study by Rodd and colleagues described earlier, increased activation in LIFC is most likely due to increased selection demands in reaction to sentences with ambiguous words. Selection is often, but not always, a prerequiste for unification. Unification with or without selection is a core feature of language processing. During natural language comprehension, information has to be kept in working memory for a certain period of time, and incoming

information has to be integrated and combined with previous information. The combinatorial nature of language necessitates that a representation be constructed online, without the availability of an existing representation of the utterance in long-term memory. In addition, some information sources that are integrated with language do not have a stable representation in long-term memory such that they can be selected. For instance, there is no stable representation of the meaning of co-speech gestures, which are highly ambiguous outside of a language context. Still, in all these cases increased activation is observed in LIFC, such as when the integration load of information from co-speech gestures is high (Willems et al., 2007). Similarly, it is unlikely that integration of information about characteristics of the speaker as indicated by the acoustics of the voice (e.g., whether the speaker is male or female, child or adult) relies on selection. Nevertheless, increased activation levels are observed in LIFC when integrating speaker characteristics with the content of the message gets more difficult (Tesink et al., in press). Therefore, unification is a more general account of LIFC function. It implies selection, but it covers additional integration processes as well.

Integration Versus Unification We have so far used the term "unification" to refer to the assembly of complex meaning. Although the term "integration" is often used as a synonym for unification, including by ourselves, we suggest that it is useful to make a functional distinction between the two. Semantic integration is at stake if different sources of information converge on a common memory representation. An example is the sound and the sight of an animal (e.g., a barking dog). The sight of a dog, the barking sound, and their combined occurrence most likely all activate a memory representation of "dog" that has multimodal characteristics. Semantic unification, however, is always a constructive process in which a semantic representation is constructed that is not already available in memory. This distinction makes opposite predictions for the BOLD response. Semantic unification is always harder for semantic incongruities. These should result in a stronger BOLD response than semantically congruent items. In contrast, congruent input results in converging support for a prestored representation, which might thus be more strongly activated compared to a situation with incongruent input. Hence, in the case of integration, the congruent condition will elicit a stronger BOLD response than the incongruent condition. A few studies on multimodal integration have indeed reported activation increases to matching stimulus combinations. For instance, Van Atteveldt, Formisano, Goebel, and Blomert (2004) observed a higher activation level in left superior temporal cortex in response to a matching phoneme and letter combination (e.g., letter "p" with phoneme [p]) as compared to a mismatching combination (e.g., letter "k" with phoneme [p]) (see also Calvert, Campbell, & Brammer, 2000, for the integration of lip movements and speech sounds). The same is true in the study by Beauchamp, Lee, Argall, and Martin (2004), who found higher activation in left posterior temporal cortex to the matching combination of a picture of an object and its sound versus an incongruent combination. In a recent paper Hein and colleagues (2007) reported an interesting difference between inferior frontal cortex (IFC) and posterior temporal cortex. The IFC showed a stronger response to incongruent familiar animal sounds and images (e.g., a meowing dog) than to the familiar combination (a barking dog). This was, however, not observed in STG and pSTS. This region was found to be more strongly activated to highly familiar combinations of objects and sounds as compared to combinations of artificial objects and sounds. This result suggests a possible division of labor between inferior frontal and superior temporal areas, with a stronger contribution to integration for temporal cortex and a stronger role for the IFC in unification—that is, in constructing a common representation that is not already available in long-term memory.

However, as we have seen, many studies on sentence processing have found increased activation, especially in left superior/middle temporal cortex when the (semantic) unification load of a word increases given the preceding sentence context (e.g., Bookheimer, 2002; Friederici et al., 2003; Kuperberg et al., 2003; Hagoort et al., 2004; Rodd et al., 2005; Ruschemeyer, Fiebach, Kempe, & Friederici, 2005; Davis et al., 2007; Willems et al., 2007, 2008). We propose that this results from signals from LIFC, indicating that in the service of unification, lexical-semantic information needs to be maintained active longer or needs to be reaccessed when unification load increases (cf. Humphries, Binder, Medler, & Liebenthal, 2007). In this way, it is the dynamic interplay between LIFC and left superior/middle temporal cortex that is necessary for successful semantic unification.

Conclusion

Over and above the retrieval of individual word meanings, sentence and discourse processing requires combinatorial operations that result in a coherent interpretation of multiword utterances. These operations do not adhere to a simple principle of compositionality. World knowledge, information about the speaker, co-occurring visual input, and discourse information all trigger electrophysiological responses similar to those triggered by sentence-internal semantic information. A network of brain areas, including the left inferior frontal gyrus, the left superior/middle temporal cortex, the left inferior parietal cortex, and, to a lesser extent, their right-hemisphere homologues are recruited to perform semantic unification. In line with the MUC framework,

semantic unification operations are under top-down control of left, and in the case of discourse, also right inferior frontal cortex. This contribution modulates activations of lexical information in memory as represented by the left superior and middle temporal cortex, presumably with additional support for unification operations in left inferior parietal areas (e.g., angular gyrus). A more precise account of the individual contributions of these core nodes in the unification network awaits further research.

ACKNOWLEDGMENTS We thank Jos van Berkum, Karl-Magnus Petersson, and the Neurocognition of Language Ph.D.s for their comments on an earlier version of this chapter.

REFERENCES

Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*, 907–918.

Baggio, G., & van Lambalgen, M. (2007). The processing consequences of the imperfective paradox. *J. Semantics*, *24*, 307–330.

Baggio, G., van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *J. Mem. Lang.*, *59*, 36–53.

Baggio, G., van Lambalgen, M., & Hagoort, P. (in press). The processing consequences of compositionality. In W. Hinzen, E. Machery, & M. Werning (Eds.), *The Oxford handbook of compositionality*. Oxford, UK: Oxford University Press.

Baumgaertner, C., Weiller, C., & Buchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *NeuroImage*, *16*, 736–745.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*, 809–823.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.*, *25*, 151–188.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *J. Cogn. Neurosci.*, *5*, 34–44.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Ann. NY Acad. Sci.*, *1124*, 1–38.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.*, *10*, 649–657.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Q. J. Exp. Psychol. [A]*, *49*, 639–663.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Lang. Cogn. Process.*, *13*, 21–58.

Culicover, P., & Jackendoff, R. (2005). *Simpler syntax*. Oxford, UK: Oxford University Press.

Cutler, A., & Clifton, C. E. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown and P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford, UK: Oxford University Press.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, *26*, 55–67.

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., Owen, A. M., & Menon, D. K. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proc. Natl. Acad. Sci. USA*, *104*, 16032–16037.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.*, *8*, 1117–1121.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 491–505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *J. Mem. Lang.*, *41*, 469–495.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.*, *33*, 182–196.

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Hum. Brain Mapp.*, *29*, 581–593.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*, 109–128.

Fodor, J., & Lepore, E. (2002). *The compositionality papers*. Oxford, UK: Oxford University Press.

Frege, G. (1892). Uber Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, *100*, 25–50.

Friederici, A. D., Ruschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cereb. Cortex*, *13*, 170–177.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*, 531–534.

Gold, B. T., Balota, D. A., Kirchhoff, B. A., & Buckner, R. L. (2005). Common and dissociable activation patterns associated with controlled semantic and phonological processing: Evidence from fMRI adaptation. *Cereb. Cortex*, *15*, 1438–1450.

Guillem, F., Rougier, A., & Claverie, B. (1999). Short- and long-delay intracranial ERP repetition effects dissociate memory systems in the human brain. *J. Cogn. Neurosci.*, *11*, 437–458.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends Cogn. Sci.*, *9*, 416–423.

Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–81). Hillsdale, NJ: Lawrence Erlbaum.

Hagoort, P., Brown, C., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown and P. Hagoort (Eds.), *The neurocognition of language* (pp. 273–317). Oxford, UK: Oxford University Press.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–441.

Halgren, E., Baudena, P., Heit, G., Clarke, J. M., Marinkovic, K., & Chauvel, P. (1994). Spatio-temporal stages in face and word processing. 2. Depthrecorded potentials in the

human frontal and Rolandic cortices. *J. Physiol. Paris*, *88*, 51–80.

Halgren, E., Dhond, R. P., Christensen, N., Petten, C. V., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage*, *17*, 1101–1116.

Hasson, U., Nusbaum, H. C., & Small, S. L. (2007). Brain networks subserving the extraction of sentence information and its encoding to memory. *Cereb. Cortex*, *17*, 2899–2913.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. New York: Blackwell.

Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.*, *27*, 7881–7887.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain*, *121*, 1133–1142.

Hoenig, K., & Scheef, L. (2005). Mediotemporal contributions to semantic processing: fMRI evidence from ambiguity processing during semantic context verification. *Hippocampus*, *15*, 597–609.

Holcomb, P. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*, 47–61.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: An fMRI study. *NeuroImage*, *36*, 924–932.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jackendoff, R. (1999). The representational structures of the language faculty and their interactions. In C. M. Brown and P. Hagoort (Eds.), *The neurocognition of language* (pp. 37–79). Oxford, UK: Oxford University Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.

Keenan, E. L. (1979). On surface form and logical form. *Stud. Linguist. Sci.*, *8*, 163–203.

Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Reading anomalous sentences: An event-related fMRI study of semantic processing. *NeuroImage*, *17*, 842–850.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *J. Mem. Lang.*, *52*, 205–225.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Res.*, *1146*, 23–49.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *J. Cogn. Neurosci.*, *15*, 272–293.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. *NeuroImage*, *33*, 343–361.

Kuperberg, G. R., McGuire, P. K., Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Wright, I. C., Lythgoe, D. J., Williams, S. C., & David, A. S. (2000). Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: An fMRI study. *J. Cogn. Neurosci.*, *12*, 321–341.

Kuperberg, G. R., Sitnikova, T., & Lakshmanan, B. M. (2008). Neuroanatomical distinctions within the semantic system during sentence comprehension: Evidence from functional magnetic resonance imaging. *NeuroImage*, *40*, 367–388.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.*, *4*, 463–470.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science*, *207*, 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.

Kutas, M., Van Petten, C., & Kluender, K. R. (2006). Psycholinguistics electrified. II. 1994–2005. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). Amsterdam: Elsevier.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford, UK: Oxford University Press.

Li, X., Hagoort, P., & Yang, Y. (2008). Event related potential evidence on the influence of accentuation in spoken discourse comprehension in Chinese. *J. Cogn. Neurosci.*, *20*, 906–915.

Maguire, E. A., Frith, C. D., & Morris, R. G. (1999). The functional neuroanatomy of comprehension and memory: The importance of prior knowledge. *Brain*, *122*, 1839–1850.

Mason, R. A., & Just, M. A. (2006). Neuroimaging contributions to the understanding of discourse processes. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 765–799). Amsterdam: Elsevier.

McCarthy, G., Nobre, A. C., Bentin, S., & Spencer, D. D. (1995). Language-related field potentials in the anterior-medial temporal lobe. I. Intracranial distribution and neural generators. *J. Neurosci.*, *15*, 1080–1089.

McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, *78*, B17–25.

Montague, R. (1970). Universal grammar. *Theoria*, *36*, 373–398.

Moss, H. E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., & Tyler, L. K. (2005). Selecting among competing alternatives: Selection and retrieval in the left inferior frontal gyrus. *Cereb. Cortex*, *15*, 1723–1735.

Newman, A. J., Pancheva, R., Ozawa, K., Neville, H. J., & Ullman, M. T. (2001). An event-related fMRI study of syntactic and semantic violations. *J. Psycholinguist. Res.*, *30*, 339–364.

Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., Shaywitz, B. A., & Gore, J. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *J. Cogn. Neurosci.*, *12*, 120–133.

Nieuwland, M. S., & van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *J. Cogn. Neurosci.*, *18*, 1098–1111.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.*, *31*, 785–806.

Osterhout, L., Kim, A., & Kuperberg, G. R. (2007). The neurobiology of sentence comprehension. In M. Spivey, M. Joanaisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics*. Cambridge, UK: Cambridge University Press.

Osterhout, L., McLaughlin, J., & Bersick, M. (1997). Event-related brain potentials and human language. *Trends Cogn. Sci.*, *1*, 203–209.

Partee, B. H. (1984). Compositionality. In F. Veltman & F. Landmand (Eds.), *Varieties of formal semantics*. Dordrecht: Foris.

Partee, B. H., Ter Meulen, A., & Wall, R. E. (1990). *Mathematical methods in linguistics*. Dordrecht: Kluwer.

Pylkkänen, L., & McElree, B. (2007). An MEG study of silent meaning. *J. Cogn. Neurosci.*, *19*, 1905–1921.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanism of speech comprehension: fMRI studies of semantic ambiguity. *Cereb. Cortex*, *15*, 1261–1269.

Ruschemeyer, S. A., Fiebach, C. J., Kempe, V., & Friederici, A. D. (2005). Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Hum. Brain Mapp.*, *25*, 266–286.

Ruschemeyer, S. A., Zysset, S., & Friederici, A. D. (2006). Native and non-native reading of sentences: An fMRI experiment. *NeuroImage*, *31*, 354–365.

Seuren, P. A. M. (1998). *Western linguistics: An historical introduction*. Oxford, UK: Blackwell.

Sieborger, F. T., Ferstl, E. C., & von Cramon, D. Y. (2007). Making sense of nonsense: An fMRI study of task induced inference processes during discourse comprehension. *Brain Res.*, *1166*, 77–91.

St George, M., Kutas, M., Martinez, A., & Sereno, M. I. (1999). Semantic integration in reading: Engagement of the right hemisphere during discourse processing. *Brain*, *122*, 1317–1325.

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*, 477–488.

Tesink, C. M. J. Y., Petersson, K. M., van Berkum, J. J. A., van den Brink, D., Buitelaar, J. K., & Hagoort, P. (in press). Unification of speaker and meaning in language comprehension: An fMRI study. *J. Cogn. Neurosci.*

Thompson-Schill, S. L., Bedny, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Curr. Opin. Neurobiol.*, *15*, 219–224.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc. Natl. Acad. Sci. USA*, *94*, 14792–14797.

Traxler, M., McElree, B., Williams, R., & Pickering, M. (2005). Context effects in coercion: Evidence from eye movements. *J. Mem. Lang.*, *53*, 1–25.

Traxler, M., Pickering, M., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye movements and self-paced reading. *J. Mem. Lang.*, *47*, 530–547.

Van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb. Cortex*, *17*, 962–974.

Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*, 271–282.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *J. Cogn. Neurosci.*, *11*, 657–671.

van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.*, *20*, 580–591.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006) Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*, 1414–1432.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*, 329–338.

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cereb. Cortex*, *17*, 2322–2333.

Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: Event-related potential and functional magnetic resonance imaging evidence of word versus picture integration into a sentence context. *J. Cogn. Neurosci.*, *20*, 1235–1249.

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *NeuroImage*, *25*, 1002–1015.

Zempleni, M. Z., Renken, R., Hoeks, J. C., Hoogduin, J. M., & Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *NeuroImage*, *34*, 1270–1279.

57 Early Language Acquisition: Neural Substrates and Theoretical Models

PATRICIA K. KUHL

ABSTRACT Infants learn language(s) with apparent ease, and the tools of modern neuroscience are providing valuable information about the mechanisms that underlie this capacity. Noninvasive, safe brain technologies have now been proven feasible for use with children starting at birth, and studies in the past decade at the phonetic, word, and sentence levels have produced an explosion in neuroscience research examining young children's language processing. At all levels of language, the neural signatures of learning can be documented at remarkably early points in development. Importantly both for theory and for the eventual application of this work to the diagnosis and treatment of developmental disabilities, early brain measures of infants' responses to phonetic differences are reflected in infants' language abilities in the second and third year of life. Developmental neuroscience studies using language are beginning to answer questions about the origins and development of human's language faculty.

Infants begin life with the capacity to detect phonetic distinctions across all languages, and they develop a language-specific phonetic capacity and acquire early words before the end of the first year (Jusczyk, 1997; Kuhl, Conboy, Padden, Rivera-Gaxiola, & Nelson, 2008; Werker & Curtin, 2005). A major question remains, however: Do infants' initial capacities and their ability to learn effortlessly from exposure to language reflect domain-specific mechanisms that operate exclusively on linguistic data or mechanisms that operate on more general learning mechanisms? In a classic debate, a nativist and a learning theorist took very different positions regarding the innate state and the nature of learning regarding language. Noam Chomsky (1959) argued that infants' innate capacities and the manner in which language was acquired were unique to language and to humans, while B. F. Skinner (1957) asserted that neither the initial state nor the manner in which language was learned was unique.

The tools of modern developmental neuroscience are bringing us closer to addressing these issues and may one day help resolve the classic debate about the interaction between biology and culture that produces the human capacity for language. Neuroscientific studies will also provide valuable information that may allow us to diagnose developmental disabilities at a stage in development when interventions are more likely to improve children's lives.

Remarkable progress has been made in the last decade in scientists' abilities to examine the young infant brain while its owner processes language, reacts to social stimuli such as faces, listens to music, or hears the voice of the child's mother. This review focuses on the new techniques and what they are teaching us about the earliest phases of language acquisition.

Neuroscientific studies on infants and young children now extend from phonemes to words to sentences. These studies fuel the hope that an understanding of development in typically developing children and in children with developmental disabilities will be achieved. Studies show that exposure to language in the first year of life begins to set the neural architecture in a way that vaults the infant forward in the acquisition of language. The goal in this chapter is to explore what we have learned about the neural mechanisms that underlie language in typically developing children, and how they differ in children with developmental disabilities that involve language such as autism.

Neuroscience techniques measure language processing in the young brain

Rapid advances have been made in the development of noninvasive techniques to examine language processing in infants and young children (figure 57.1). These methods include electroencephalography (EEG)/event-related potentials (ERPs), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and near-infrared spectroscopy (NIRS). ERPs have been widely used to study speech and language processing in infants and young children (for reviews see Conboy, Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2008; Friederici, 2005; Kuhl, 2004; Kuhl

PATRICIA K. KUHL Institute for Learning and Brain Sciences, University of Washington, Seattle, Washington

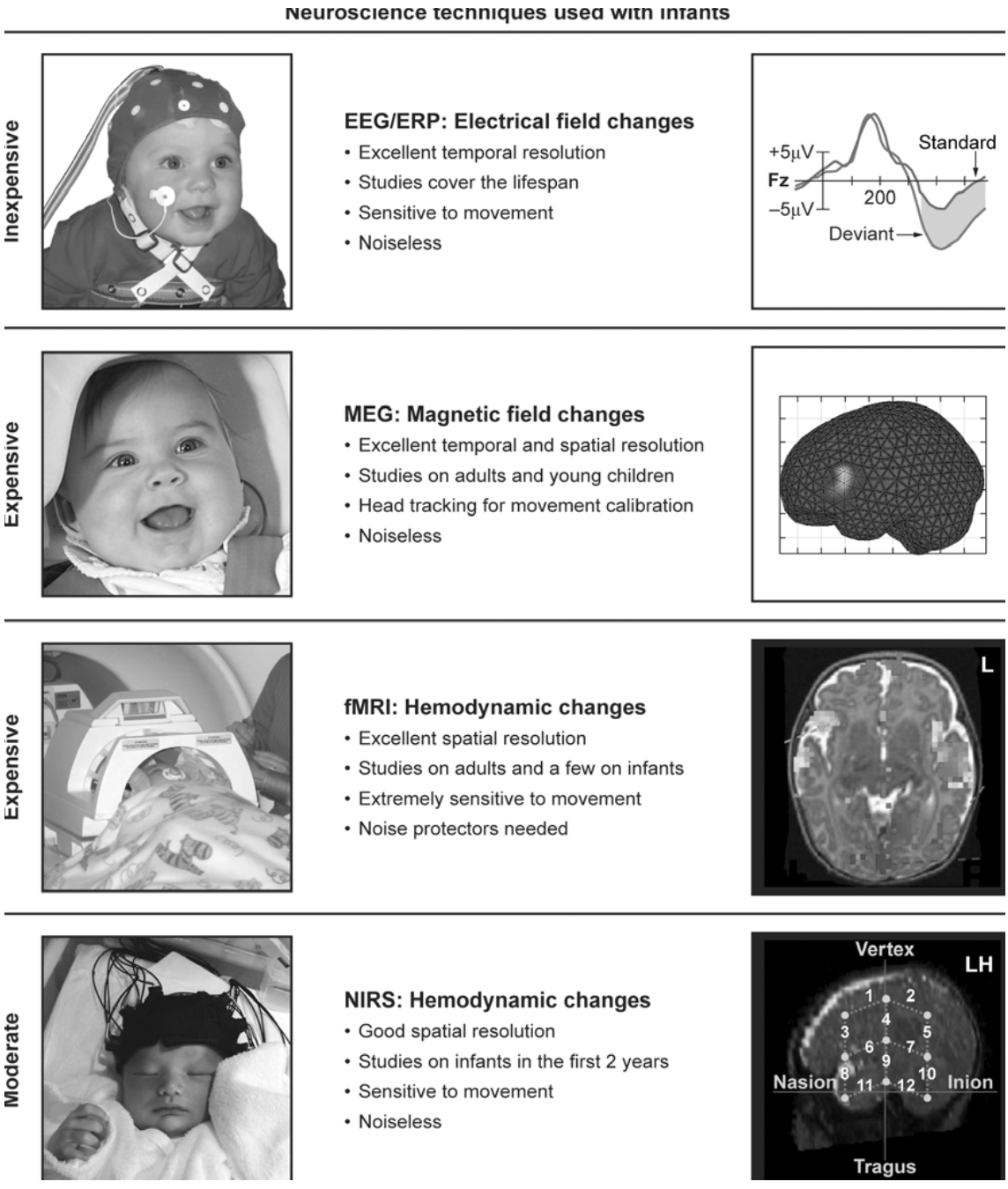


FIGURE 57.1 Four neuroscience techniques now used with infants and young children to examine the brain's responses to linguistic signals. (From Kuhl & Rivera-Gaxiola, 2008.)

& Rivera-Gaxiola, 2008). Event-related potentials (ERPs), a part of the EEG, reflect electrical activity that is time-locked to the presentation of a specific sensory stimulus (e.g., syllables, words) or a cognitive process (recognition of a semantic violation within a sentence or phrase). By placing sensors on a child's scalp, the activity of neural networks firing in a coordinated and synchronous fashion in open field configurations can be measured, and voltage changes occurring as a function of cortical neural activity can be detected. ERPs provide precise time resolution (milliseconds), making them well suited for studying the high-speed and temporally ordered structure of human speech. ERP experiments can also be carried out in populations who, because of age or cognitive impairment, cannot provide overt responses. Spatial resolution of the source of brain activation is, however, limited.

Magnetoencephalography (MEG) is another brain-imaging technique that tracks activity in the brain with

exquisite temporal resolution. MEG (as well as EEG) techniques are safe and and noiseless, allowing data collection while infants listen to language in a quiet environment. The SQUID (superconducting quantum interference device) sensors located within the MEG helmet measure the minute magnetic fields associated with electrical currents that are produced by the brain when it is performing sensory, motor, or cognitive tasks. MEG allows precise localization of the neural currents responsible for the sources of the magnetic fields, and it has been used to test phonetic discrimination in adults (Kujala, Alho, Service, Ilmoniemi, & Connolly, 2004).

Recently a genuine advance was documented by the first MEG studies testing awake infants in the first year of life (Bosseler et al., 2008; Cheour et al., 2004; Imada et al., 2006, 2008). In these studies, the use of sophisticated head-tracking software and hardware allows correction for infants' head movements, so infants are free to move comfortably during the tests. MEG studies allow whole-brain imaging during speech discrimination, which is now providing data on the location and timing of brain activation in critical regions (Broca's and Wernicke's) involved in language acquisition (see Bosseler et al.; Imada et al., 2006, 2008).

MEG and/or EEG can be combined with magnetic resonance imaging (MRI), a technique that provides static structural/anatomical pictures of the brain. Using mathematical modeling methods, the specific brain regions that produce the magnetic or electrical signals can be identified in the human brain with high spatial resolution (millimeter). Structural MRIs allow measurement of anatomical changes in white and gray matter in specific brain regions across the life span. MRIs can be superimposed on the physiological activity detected by MEG or EEG to refine the spatial localization of brain activities for individual participants.

Functional magnetic resonance imaging (fMRI) is now considered a standard method of neuroimaging in adults because it provides high-spatial-resolution maps of neural activity across the entire brain (e.g., Gernsbacher & Kachak, 2003). However, unlike EEG and MEG, fMRI does not directly detect neural activity, but rather the changes in blood oxygenation that occur in response to neural activation/firing. Neural events happen in milliseconds, while the blood-oxygenation changes that they induce are spread out over several seconds, thereby severely limiting fMRI's temporal resolution. Adult studies are employing new fMRI data-analysis methods for speech stimuli and correlating the fMRI data to behavioral data. For example, Raizada, Tsao, Liu, and Kuhl (2009), using a multivariate pattern classifier, showed that English—but not Japanese—speakers exhibited distinct neural activity patterns for /ra/ and /la/ in primary auditory cortex. Subjects who behaviorally distinguished the sounds most accurately also had the most distinct neural activity patterns.

Functional MRI techniques would be very valuable with infants, but few studies have attempted fMRI with infants (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Dehaene-Lambertz, Hertz-Pannier, Dubois, Meriaux, & Roche, 2006). The technique requires subjects to be perfectly still, and the MRI device produces loud sounds making it necessary to shield infants' ears while delivering language stimuli.

Near-infrared spectroscopy (NIRS) also measures cerebral hemodynamic responses in relation to neural activity, but employs the absorption of light, which is sensitive to the concentration of hemoglobin, to measure activation (Aslin & Mehler, 2005). NIRS utilizes near-infrared light to measure changes in blood oxy- and deoxyhemoglobin concentrations in the brain as well as total blood volume changes in various regions of the cerebral cortex. The NIRS system can determine where and how active the specific regions of the brain are by continuously monitoring blood hemoglobin levels, and reports have begun to appear on infants in the first two years of life (Bortfeld, Wruck, & Boas, 2007; Homae, Watanabe, Nakano, Asakawa, & Taga, 2006; Pena et al., 2003; Taga & Asakawa, 2007). Homae and colleagues, for example, provided data using NIRS that suggest that sleeping 3-month-old infants process the prosodic information in sentences in the right temporoparietal region. As with other techniques relying on hemodynamic changes such as fMRI, NIRS does not provide good temporal resolution. One of the most important uses of this technique is that coregistration with other testing techniques such as EEG and MEG may be possible.

The use of these techniques with infants and young children has produced an explosion of neuroscience studies using stimuli that tap all levels of language—phoneme, word, and sentence. In the next sections, examples of recent findings will be described to give a sense of the promise of neuroscience for the study of language acquisition in children.

Neural signatures of phonetic learning in typically developing children

Perception of the basic units of speech—the vowels and consonants that make up words—is one of the most widely studied behaviors in infancy and adulthood, and studies using ERPs have advanced our knowledge of development and learning.

Behavioral studies demonstrated that at birth young infants exhibit a universal capacity to detect differences between phonetic contrasts used in the world's languages (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). We have referred to this as Phase 1 in development (Kuhl et al., 2008). This universal capacity is dramatically altered by language experience starting as early as 6 months for vowels and by 10 months for consonants: over time, native language

phonetic abilities significantly increase (Cheour et al., 1998; Kuhl et al., 2006; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005; Sundara, Polka, & Genesee, 2006) while the ability to discriminate phonetic contrasts that are not relevant to the language of the culture declines (Best & McRoberts, 2003; Cheour et al., 1998; Kuhl et al., 2006; Rivera-Gaxiola, Silva-Pereyra, et al., 2005; Werker & Tees, 1984).

By the end of the first year, the infant brain is no longer universally prepared for all languages, but primed to acquire the specific one(s) to which they have been exposed. We refer to this as Phase 2 in infant phonetic development (Kuhl et al., 2008). The explanation of this transition from Phase 1 to Phase 2 has become the focus of intense study because it illustrates the interaction between biology and culture—between infants' initial state and infants' abilities to learn. Speech offers the opportunity to study the brain's ability to be shaped implicitly by experience.

Kuhl and colleagues (2008) examined whether the transition in phonetic perception from a language-general ability to a language-specific one—from Phase 1 to Phase 2—can be linked to the growth of language. The work provided a critical test stemming from the native language neural commitment (NLNC) hypothesis (Kuhl, 2004). According to NLNC, initial native language learning involves *neural commitment* to the patterned regularities contained in ambient speech, with bidirectional effects: neural coding facilitates the detection of more complex language units (words) that build on initial learning, while simultaneously reducing attention to alternate patterns, such as those of a foreign language.

This formulation suggests that infants with excellent phonetic learning skills should advance more quickly toward language. In contrast, foreign-language phonetic perception reflects the degree to which the infant brain remains uncommitted to native-language patterns—still in Phase 1 as it were—at a more universal and immature phase of development. Infants in Phase 1 remain "open" to nonnative-speech patterns. As an open system reflects uncommitted circuitry, infants who remain highly skilled at discriminating foreign-language phonetic units would be expected to show a slower progression toward language.

New ERP studies of infants support the NLNC assertion. Kuhl and colleagues (2008) measured infants' ERPs at 7.5 months of age in response to changes in native (/p-t/) and nonnative (Mandarin /ɕ-tɕ^{h}/ and Spanish /t-d/) phonemes. The mismatch negativity (MMN), which has been shown in adults to be a neural correlate of phonetic discrimination (Naatanen et al., 1997), was calculated for both the native and nonnative phonemes for each infant. Individual variation was observed for both native and nonnative discrimination, representing either "noise" or meaningful differences among infants.

The results supported the idea that the differences among infants were meaningful. MMN measurements taken at 7.5 months—for both the native and the nonnative phonetic contrasts—predicted later language. However, and in accord with the NLNC hypothesis, the native and nonnative contrasts predicted language growth in opposing directions (Kuhl et al., 2008).

The MMN component was elicited in individual infants (figure 57.2*A*). Native and nonnative contrasts were measured in counterbalanced order, and the MMN was observed between 250 and 400 ms (figure 57.2*B*). For the infant shown in figure 57.2*A*, greater negativity of the MMN, indicating better neural discrimination, was shown for the native when compared to the nonnative phonetic contrast; other infants showed equal discrimination for the two contrasts or better discrimination of the nonnative contrast. Infants' language abilities were measured at four later points in time: 14, 18, 24, and 30 months of age using the MacArthur-Bates Communicative Development Inventories (CDI), a reliable and valid measure assessing language and communication development from 8 to 30 months of age (Fenson et al., 1993).

The MMN measures taken at 7.5 months of age were related to the language measures taken between 14 and 30 months of age. For the native contrast, the strength of the MMN (better discrimination) predicted accelerated word production at 24 months, greater sentence complexity at 24 months, and longer mean length of utterance at 30 months of age. In contrast, for the nonnative stimulus pair, the strength of the MMN at the same age in the same infants predicted slower language development at the same future points in time. Behavioral (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005) and brain measures (Kuhl et al., 2008), collected on the same infants, were significantly correlated.

This pattern, showing differential effects of good discrimination for the native and nonnative contrasts, can be readily seen in the growth of vocabulary from 14 to 30 months (figure 57.2*C*). Hierarchical Linear Growth Curve modeling (Raudenbush, Bryk, Cheong, & Congdon, 2005) shows that both native and nonnative discrimination at 7.5 months significantly predicts vocabulary growth, but the effects of good phonetic discrimination are reversed for the native and nonnative predictors. Better native phonetic discrimination predicts accelerated vocabulary growth, whereas better nonnative phonetic discrimination predicts slower vocabulary growth (Kuhl et al., 2008). These results support the NLNC hypothesis.

Rivera-Gaxiola and colleagues (Rivera-Gaxiola, Klarman, Garcia-Sierra, & Kuhl, 2005; Rivera-Gaxiola, Silva-Pereyra, et al., 2005) demonstrated a similar pattern of prediction using a different nonnative contrast. They recorded auditory ERP complexes in 7- and 11-month-old American infants in response to both Spanish and English voicing contrasts.

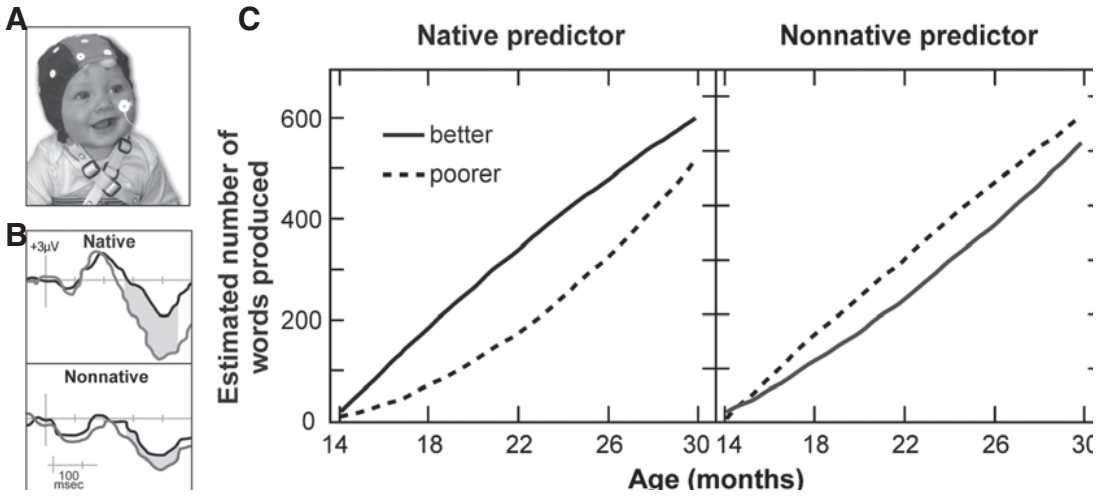


FIGURE 57.2 (*A*) A 7.5-month-old infant wearing an ERP electrocap. (*B*) Infant ERP waveforms at one sensor location (CZ) for one infant are shown in response to native (English) and nonnative (Mandarin) phonetic contrast at 7.5 months. The mismatch negativity (MMN) is obtained by subtracting the standard waveform (black) from the deviant waveform (gray). This infant's response suggests that native-language learning has begun because the MMN negativity in response to the native English contrast is considerably stronger (more negative) than that to the nonnative constrast. (*C*) Hierarchical linear growth modeling of vocabulary growth between 14 and 30 months is shown for two groups of children, those whose MMN values at 7.5 months indicated better discrimination (−1 SD) and those MMN values indicated poorer discrimination (+1 SD). Vocabulary growth was significantly faster for infants with better MMN phonetic discrimination for the native contrast at 7.5 months of age (*C, left*). In contrast, infants with better discrimination for the nonnative contrasts (−1 SD) as indicated by MMN at 7.5 months showed slower vocabulary growth (*C, right*). Both contrasts predict vocabulary growth, but the effects of better discrimination are reversed for the native and nonnative contrasts. (From Kuhl & Rivera-Gaxiola, 2008.)

Two patterns of ERP response were observed—an early positive-going wave (P150–250) and a later negative-going wave (N250–550) (Rivera-Gaxiola, Silva-Pereyra, et al., 2005). Further work examined the patterns of the same auditory ERP positive-negative complexes in a larger sample of 11-month-old monolingual American infants using the same contrasts used in the developmental study, and found that infants' response to the nonnative contrast predicted the number of words produced at 18, 22, 25, 27, and 30 months of age (Rivera-Gaxiola, Klarman, et al., 2005). Infants showing an N250–550 to the foreign contrast at 11 months of age (indexing better neural discrimination) produced significantly fewer words at all ages when compared to infants showing a less negative response. Scalp distribution analyses on 7-, 11-, 15-, and 20-month-old infants revealed that the P150–250 and the N250–550 components differ in distribution (Rivera-Gaxiola et al., 2007). Thus in both Kuhl and colleagues (2008) and Rivera-Gaxiola, Klarman, and colleagues (2005), an enhanced negativity in response to the nonnative contrast is associated with slower language development.

The continuity in language development documented in these studies using infants' early phonetic skills to predict later language (Kuhl, Conboy, et al., 2005; Kuhl et al., 2008; Rivera-Gaxiola, Klarman, et al., 2005; Tsao, Liu, & Kuhl, 2004) is also seen in studies that use infants' early pattern detection skills for speech to predict later language (Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006), as well as in studies that use infants' early processing efficiency for words to predict later language (Fernald, Perfors, & Marchman, 2006). Taken as a whole, these studies form bridges between the early precursors to language in infancy and measures of language competencies in early childhood, bridges that are important to theory building as well as to clinical populations with developmental disabilities that involve language.

ERP studies at the phonetic level suggest that the young brain's response to the elementary building blocks of language matters and that initial native language phonetic learning is a pathway to language (Kuhl, 2008). The data also suggest that discriminating nonnative phonetic contrasts for a longer period of time in early development—reflecting infants' initial, more immature state—can be linked to slower language development. In infants exposed to a single language, the ability to attend to changes in the phonetic contrasts that are relevant to the culture's language, while at the same time reducing attention to phonetic contrasts from other languages that are discriminable but irrelevant to the language of their culture, appears to be an important first step toward the acquisition of language. What neuroscience tools may allow us to do in the future is to understand this process and its relation to the "critical period" for language development (see Kuhl, Conboy, et al., 2005, for discussion).

Brain measures of learning from exposure to a second language

Recent studies have shown that young infants are capable of phonetic learning at 9 months of age from exposure to a new language but only when exposure occurs during live human presentation; television or audio-only exposure did not produce learning (Kuhl, Tsao, & Liu, 2003). Social interaction appears to be a critical component for language learning, a finding that ties early communicative learning in speech to examples of communicative learning in neurobiology more generally, as shown by the importance of social factors in song learning in birds (e.g., Brainard & Knudsen, 1998). I have used these second-language exposure studies to argue that the social brain may "gate" the computational mechanisms underlying language learning during the earliest stages of human language acquisition (Kuhl, 2007).

The social "gating" hypothesis was tested in studies using ERP measures of second-langauge learning (Conboy, Brooks, Taylor, Meltzoff, & Kuhl, 2008; Conboy & Kuhl, 2007). In the study, American monolingual infants were exposed to Spanish at 9 months of age by native Spanish speakers, and their ability to learn both phonemes and words from this foreign-language exposure was tested. The study tested the social hypothesis by examining whether the infants' tendency to interact in socially sophisticated ways during the exposure sessions would predict the degree to which individual infants learned both phonemes and words from the new language.

Infants' ERPs in response to English and Spanish phonemes, as well as their ERP responses to Spanish words, were measured before and after exposure to Spanish. As in the Mandarin study, exposure consisted of live interaction with foreign-language "tutors" during 12 sessions, each of which lasted 25 minutes. All sessions were videotaped using a four-camera system, and detailed measures of shared visual attention between the infants and their tutors were taken by an independent observer.

The ERP results demonstrated that the MMN response to the Spanish contrast was not present before exposure, but that following exposure to Spanish, the MMN was robust (Conboy & Kuhl, 2007), replicating the behavioral findings in the Mandarin study (Kuhl et al., 2003). This result confirms infants' ability to learn phonetically from exposure to a foreign language at 9 months of age. Extending these previous findings beyond phoneme learning, Conboy and Kuhl also showed that infants learned Spanish words that were presented during the exposure sessions. When compared to Spanish words that had not been presented, infants' ERPs to the Spanish words revealed the classic components related to known words (Conboy & Kuhl, 2007).

The social gating hypothesis was also strongly supported. Infants' degree of social engagement—for example, the degree to which infants alternated their visual attention between a newly presented toy and the tutor's eyes, as opposed to simply focusing on the toy or on the tutor—predicted the degree of learning both for phonemes and for words (Conboy, Brooks, et al., 2008). In other words, the degree to which an individual infant interacted socially during the 12 language sessions predicted the degree of learning measured well after the four-week exposure was complete (Conboy, Brooks, et al., 2008). Gaze following has previously been shown to predict word learning in infants (Brooks & Meltzoff, 2008). These results show that the relationship between social interaction and language learning can be demonstrated experimentally for new learning of language material at 9 months of age. Finally, the results of the study suggest the possibility that exposure to a new language provides a cognitive enhancement. Pre- and post-exposure measures of "cognitive control"—the ability to attend selectively and inhibit prepotent responses, which has previously been shown to be enhanced in bilingual adults (Bialystok, 1999) and children (Carlson & Meltzoff, 2008)—were also obtained from the children involved in the language exposure experiments. These measures indicated that cognitive control skills are enhanced after, but not prior to, Spanish exposure, linking bilingual learning to the enhancement of particular cognitive skills (Conboy, Sommerville, & Kuhl, 2008).

In sum, ERPs provide a highly sensitive measure of learning for both phonemes and words. ERP responses to speech not only predict the growth of language over the first 30 months (Kuhl et al., 2008; Rivera-Gaxiola, Klarman, et al., 2005), but are also sufficiently sensitive to reflect the subtle abilities that contribute to infant learning, such as infants' social eye-gaze following (Conboy, Brooks, et al., 2008). Complex natural language learning may demand social interaction, because language evolved in a social setting. The neurobiological mechanisms underlying language likely utilized interactional cues made available only in a social setting. In the future, whole-brain measures, such as those provided by MEG, will allow us to observe brain activation during live presentations of language versus those that are merely televised to explore hypotheses about why human interaction is essential to language learning (Kuhl et al., 2003). Moreover, using "social" robots, we are now conducting studies that will define what constitutes a social agent for a young child (Virnes, Cardillo, Kuhl, & Movellan, 2008).

Neural signatures of word learning

A sudden increase in vocabulary typically occurs between 18 and 24 months of age—a "vocabulary explosion" (Fernald et al., 2006; Ganger & Brent, 2004)—but word learning starts much earlier. Infants show recognition of their own

name at 4.5 months (Mandel, Jusczyk, & Pisoni, 1995). At 6 months, infants use their own names or the word *Mommy* in an utterance to identify word boundaries (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005) and look appropriately to pictures of their mother or father when hearing *Mommy* or *Daddy* (Tincoff & Jusczyk, 1999). By 7 months, infants listen longer to passages containing words they previously heard rather than passages containing words they have not heard (Jusczyk & Hohne, 1997), and by 11 months infants prefer to listen to words that are highly frequent in language input over infrequent words (Halle & de Boysson-Bardies, 1994).

Behavioral studies indicate that infants learn words using both "statistical learning" strategies in which the transitional probabilities between syllables are exploited to identify likely words (Newport & Aslin, 2004; Saffran, 2003; Saffran, Aslin, & Newport, 1996) and pattern detection strategies in which infants use the typical pattern of metric stress that characterizes ambient language to segment running speech into likely words (Cutler & Norris, 1988; Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Johnson & Jusczyk, 2001; Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006).

How is word recognition evidenced in the brain? ERPs in response to words index word familiarity as early as 9 months of age and word meaning by 13–17 months of age. ERP studies have shown differences in amplitude and scalp distributions for components that are related to words that are known versus unknown to the child (Mills, Coffey-Corina, & Neville, 1993, 1997; Mills, Plunkett, Prat, & Schafer, 2005; Molfese, 1990; Molfese, Morse, & Peters, 1990; Molfese, Wetzel, & Gill, 1993; Thierry, Vihman, & Roberts, 2003).

As early as 9 months of age, ERPs indicate word familiarity, and by 13–17 months of age, studies show ERP components that reliably signal the brain's coding of words that are known versus unknown by the child (Mills et al., 1993, 1997, 2005; Molfese et al., 1990, 1993; Thierry et al., 2003). Toddlers with larger vocabularies tend to have a more focalized and larger N200 for known words—they show an enhanced negativity to known versus unknown words only at left temporal and parietal electrode sites—whereas children with smaller vocabularies show more broadly distributed effects (Mills et al., 1993), features that also distinguish typically developing preschool children from preschool children with autism (Coffey-Corina, Padden, Kuhl, & Dawson, 2007).

Processing efficiency for phonemes and words can be seen as well in the relative focalization and duration of brain activation in adult MEG studies (Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005), indicating that these features index language experience and proficiency not only in children (Conboy, Rivera-Gaxiola, et al., 2008; Friederici, 2005), but also over the life span. Individual differences in the response latency to a familiar word at the age of 2 are related to both lexical and grammatical measures collected between 15 and 25 months, providing more evidence that processing speed is associated with greater language facility (Fernald et al., 2006).

Mills and colleagues (2005) used ERPs in 20-month-old toddlers to examine new word learning. The children listened to known and unknown words, and to nonwords that were phonotactically legal in English. ERPs were recorded as the children were presented with novel objects paired with the nonwords. After the learning period, ERPs to the nonwords that had been paired with novel objects were shown to be similar to those of previously known words, suggesting that that new words may be encoded in the same neural regions as previously learned words.

ERP studies on German infants reveal the development of word-segmentation strategies based on the typical stress patterns of German words. When presented with bisyllabic strings with either a trochaic (typical in German) or iambic pattern, infants who heard a trochaic pattern embedded in an iambic string showed the N200 ERP component, similar to that elicited in response to a known word, whereas infants presented with the iambic bisyllable embedded in the trochaic pattern showed no response (Weber, Hahne, Friedrich, & Friederici, 2004). The data suggest that German infants at this age are applying a metric segmentation strategy, consistent with the behavioral data of Höhle and colleagues (2009).

Infants' early lexicons

There is evidence suggesting that young children's word representations are phonetically underspecified. Children's growing lexicons must code words in a way that distinguishes words from one another, and, given that by the end of the first year infants' phonetic skills are language specific (Best & McRoberts, 2003; Kuhl et al., 2006; Werker & Tees, 1984), it was assumed that children's early word representations were phonetically detailed. However, studies suggest that learning new words taxes young children's capacities, and that as a result, new word representations are not phonetically complete.

Reactions to mispronunciations—the age at which children no longer accept *tup* for *cup* or *bog* for *dog*—provides information about phonological specificity. Studies across languages suggest that by one year of age mispronunciations of common words (Fennel & Werker, 2003; Jusczyk & Aslin, 1995), words in stressed syllables (Vihman, Nakai, DePaolis, & Halle, 2004), or monosyllabic words (Swingley, 2005) are not accepted as target words, indicating well-specified representations. Other studies using visual fixation of two targets (e.g., apple and ball) while one is named (*Where's the ball?)*

show that between 14 and 25 months children's tendencies to fixate the target item when it is mispronounced diminish over time (Bailey & Plunkett, 2002; Ballem & Plunkett 2005; Swingley & Aslin, 2000, 2002).

However, behavioral and neural evidence suggests that learning new words can tax children's phonological skills. Stager and Werker (1997) demonstrated that 14-month-old infants fail to learn new words when similar-sounding phonetic units are used to distinguish those words ("bih" and "dih"), but do learn if the two new words are distinct phonologically ("leef" and "neem"). By 17 months of age, infants can learn to associate similar-sounding nonsense words to novel objects (Bailey & Plunkett, 2002; Werker, Fennell, Corcoran, & Stager, 2002). Infants with larger vocabularies succeeded on this task even at the younger age, suggesting the possibility that infants with greater phonetic learning skills acquire new words more rapidly, consistent with studies showing that better native phonetic learning skills are associated with advanced word-learning skills (Kuhl, Conboy, et al., 2005; Kuhl et al., 2008; Rivera-Gaxiola, Klarman et al., 2005; Tsao et al., 2004).

Mills and colleagues (2004) used ERPs to corroborate these results. They compared ERP responses to familiar words that were either correctly pronounced or mispronounced, as well as nonwords. At the earliest age tested, 14 months, a negative ERP component (N200–400) distinguished known versus dissimilar nonsense words (*bear* versus *kobe*) but not known versus phonetically similar nonsense words (*bear* versus *gare*). By 20 months, this same ERP component distinguished correct pronunciations, mispronunciations, and nonwords, supporting the idea that between 14 and 20 months children's phonological representations of early words become increasingly detailed. Other evidence of early processing limitations stems from infants' failure to learn a novel word when its auditory label closely resembles a word they already know (*gall* which closely resembles *ball*), suggesting lexical competition effects (Swingley & Aslin, 2007).

How phonetic and word learning interact—and whether the progression is from phonemes to words, words to phonemes, or bidirectional—is a topic of strong interest that will be aided by the use of neuroscientific methods. Recent theoretical models of early language acquisition such as NLM-e (Kuhl et al., 2008) and PRIMER (Werker & Curtin, 2005) suggest that phonological and word learning may bidirectionally influence one another. On the one hand, infants with better phonetic learning skills advance more quickly toward language because phonetic skills assist the detection of phonotactic patterns, the detection of transitional probabilities in adjacent syllables, and the ability to phonologically distinguish minimally contrastive words (Kuhl, Conboy, et al., 2005). On the other hand, the more words children learn, the more crowded lexical space becomes, putting pressure on children to attend to the phonetic units that distinguish them (see Swingley & Aslin, 2007, for discussion). Further studies examining both phoneme and word learning in the same children, as in the studies using exposure to a foreign language and ERP measures as assessments of learning, will help address this issue (Conboy & Kuhl, 2007).

ERP research shows that the young brain has difficulty representing phonetic detail when focused on the task of assigning a new auditory label to a novel object. ERP results also show that brain signatures distinguish words that are known from ones that are unfamiliar to toddlers. ERPs recorded to words in the first two years suggest that experience with words results in the formation of neural representations of those words that are increasingly well specified toward the end of the second year of life.

Neural signatures of early sentence processing

To understand sentences, the child must have exquisite phonological abilities that allow segmentation of the speech signal into words, and the ability to extract word meaning. In addition, the relationship among words composing the sentence—between a subject, its verb, and its accompanying object—must be deciphered to arrive at a full understanding of the sentence. Human language is based on the ability to process hierarchically structured sequences (Friederici, Fiebach, Schlesewsky, Bornkessel, & von Cramon, 2006).

Electrophysiological components have been recorded in children and contribute to our knowledge of when and how the young brain decodes syntactic and semantic information in sentences. In adults, specific neural systems process semantic versus syntactic information within sentences, and the ERP components elicited in response to syntactic and semantic anomalies are well established (figure 57.3). For example, a negative ERP wave occurring between 250 and 500 ms that peaks around 400 ms, referred to as the N400, is elicited to semantically anomalous words in sentences (Kutas, 1997). A late positive wave peaking at about 600 ms and largest at parietal sites, known as the P600, is elicited in response to syntactically anomalous words in sentences (Friederici, 2002). And a negative wave over frontal sites between 300 and 500 ms, known as the "late anterior negativity" (LAN), is elicited in response to syntactic and morphological violations (Friederici, 2002).

Beginning in the second year of life, ERP data on sentence processing in children suggest that adultlike components in response to semantic and syntactic violations can be elicited, but also that there are differences in the latencies and scalp distributions of these components in children and adults (Harris, 2001; Friederich & Friederici, 2005, 2006; Oberecker & Friederici, 2006; Oberecker, Friedrich,

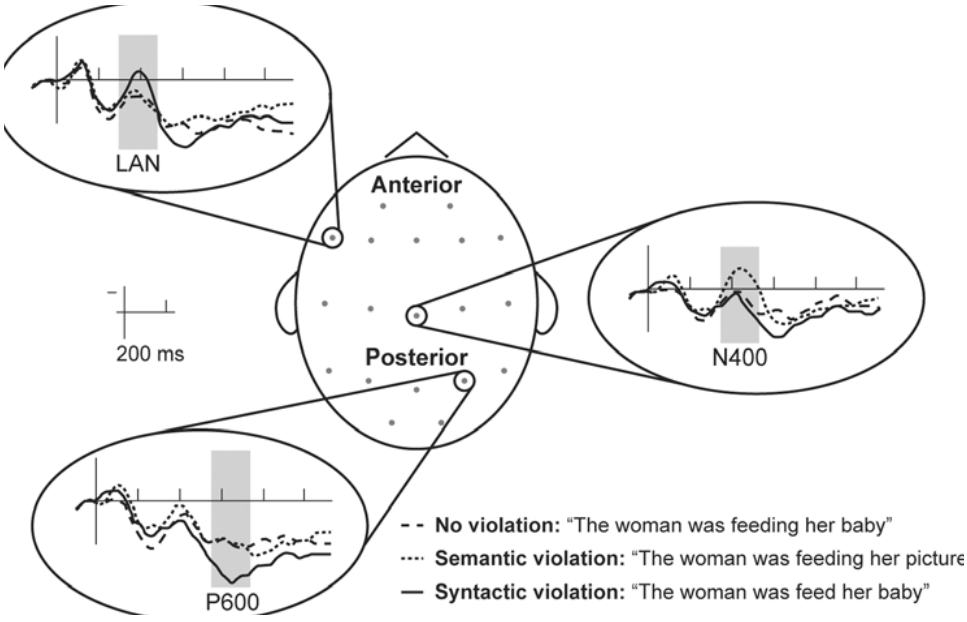


FIGURE 57.3 ERP responses to normal sentences and sentences with either semantic or syntactic anomalies show distinct distribution and polarity differences in adults. (From Kuhl & Rivera-Gaxiola, 2008.)

& Friederici, 2005; Silva-Pereyra, Conboy, Klarman, & Kuhl, 2007; Silva-Pereyra, Klarman, Lin, & Kuhl, 2005; Silva-Pereyra, Rivera-Gaxiola, & Kuhl, 2005). Holcomb, Coffey, and Neville (1992) reported the N400 in response to semantic anomaly in children from 5 years of age to adolescence; the latency of the effect was shown to decline systematically with age (see also Hahne, Eckstein, & Friederici, 2004; Neville, Coffey, Holcomb, & Tallal, 1993). Studies also show that syntactically anomalous sentences elicit the P600 in children between 7 and 13 years of age (Hahne et al.).

Recent studies have examined these ERP components in preschool children. Harris (2001) reported an N400-like effect in 36–38-month-old children, which was largest over posterior regions of both hemispheres, unlike the adult scalp distribution. Friederich and Friederici (2005) observed an N400-like wave to semantic anomalies in 19- and 24-month-old German-speaking children.

Silva-Pereyra, Rivera-Gaxiola, and Kuhl (2005) recorded ERPs in children between 36 and 48 months of age in response to semantic and syntactic anomalies. In both cases the ERP effects in children were more broadly distributed and elicited at later latencies than in adults. In work with even younger infants (30 month olds), Silva-Pereyra, Klarman, and colleagues (2005) used the same stimuli and observed late positivities distributed broadly posteriorally in response to syntactic anomalies and anterior negativities in response to semantically anomalous sentences, though in each case with longer latencies than seen in the older children and in adults (figure 57.4), a pattern seen repeatedly and attributed to the immaturities and inefficiencies of the developing processing mechanisms.

Syntactic processing of sentences with semantic content information removed—"jabberwocky sentences"—has also been tested using ERP measures with children. Silva-Pereyra and colleagues (2007) recorded ERPs to phrase structure violations in 36-month-old children using sentences in which the content words were replaced with pseudowords while leaving grammatical function words intact. The ERP components elicited to the jabberwocky phrase-structure violations differed from the same violations in real sentences. Two negative components, one from 750 to 900 ms and the other from 950 to 1050 ms, rather than the positivities seen in response to phrase structure violations in real sentences in the same children, were observed. Jabberwocky studies with adults (Canseco-Gonzalez, 2000; Hahne & Jeschenick, 2001; Munte, Matzke, & Johanes, 1997) have also reported negative-going waves for jabberwocky sentences, though at much shorter latencies.

ERP measures of early language processing in children with autism spectrum disorder (ASD)

Scientific discoveries on the progression toward language by typically developing children are now providing new insights into the language deficit shown by children with autism spectrum disorder (ASD). Neural measures of language processing in children with autism, involving both phonemes and words, when coupled with measures of ASD children's social interest in speech, are revealing a tight coupling

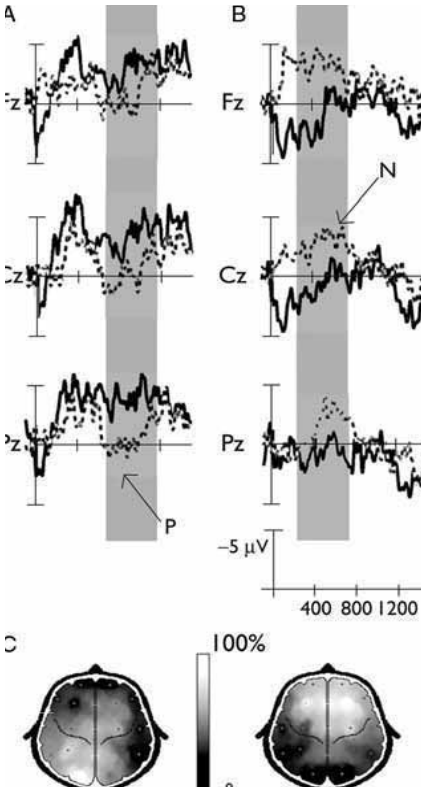


FIGURE 57.4 ERP waveforms elicited from 30-month-old children in response to sentences with (*A*) syntactic or (*B*) semantic violations. (*C*) Children's ERP responses resemble those of adults (see figure 57.3) but have longer latencies and are more broadly distributed. (From Silva-Pereyra, Klarman, Lin, & Kuhl, 2005.)

between social interaction skills and language acquisition. These measures hold promise as potential diagnostic markers of risk for autism in very young children, and therefore there is a great deal of excitement surrounding the application of these basic measures of speech processing in very young children with autism.

In typically developing children, ERP responses to simple speech syllables such as "pa" and "ta" predict the growth of language to the age of 30 months (Kuhl et al., 2008). It is therefore interesting to test whether ERP measures of autism at the phonetic level are sensitive to the degree of severity of autism, and also the degree to which the brain's responses to syllables can be predicted by other factors, such as a social interest in speech.

The first study of preschool-aged children with ASD using ERP methods examined phonetic perception (Kuhl, Coffey-Corina, Padden, & Dawson, 2005). ERPs to a simple change in two speech syllables, as well as a measure of social interest in speech, were taken. In these experiments, a listening choice test allowed young toddlers with autism to select between listening to motherese or nonspeech signals in which the formant frequencies of speech were matched by pure tones—the resulting signal was a computer warble that exactly followed the frequencies and amplitudes of the 5-s speech samples over time. Slight head turns in one direction or the other allowed the toddlers to choose their preferred signal on each trial. The goal was to compare performance at the group level between typically developing children and children with ASD, as well as to examine the relationship between brain measures of speech perception and measures of social processing of speech in children with ASD.

Considering first the ERP measures of phonetic perception, the results showed that, as a group, children with ASD exhibited no MMN to the simple change in syllables. However, when children with ASD were subgrouped on the basis of their preference for infant-directed (ID) speech (often called motherese), very different results were obtained.

The results showed that while typically developing children listen to both signals, children with autism strongly preferred the nonspeech-analogue signals. Moreover, the degree to which they did so was significantly correlated with both the severity of autism symptoms, and individual children's MMN responses to speech syllables. Toddlers with ASD who preferred motherese produced MMN responses that resembled those of typically developing children, whereas those who preferred the nonspeech analogue did not show an MMN response to the change in a speech syllable.

These results underscore the importance of a social interest in speech early in development, especially an interest in motherese. Research has shown that the phonetic units in motherese are acoustically exaggerated, making them more distinct from one another (Burnham, Kitamura, & Vollmer-Conna, 2002; Englund, 2005; Kuhl et al., 1997; Liu, Kuhl, & Tsao, 2003; Liu, Tsao, & Kuhl, 2007). Infants whose mothers use the exaggerated phonetic patterns to a greater extent when talking to them show significantly better performance in phonetic discrimination tasks (Liu et al., 2003). In the absence of a listening preference for motherese, children with autism would miss the benefit these exaggerated phonetic cues provide.

Infant-directed speech also produces unique brain responses in typically developing infants. Brain measures of typical infants' response to infant-directed speech, used by Pena and colleagues (2003) in the first study using NIRS, showed more activation in left temporal areas when infants were presented with infant-directed speech as opposed to backward speech or silence. Bortfeld and colleagues (2007) obtained analogous results using NIRS in a sample of 6–9-month-old infants presented with infant-directed speech and visual stimulation. It will be of interest to examine brain activation while children with autism listen to motherese as opposed to acoustically matched nonspeech signals. In children with ASD, brain activation to carefully controlled

speech versus nonspeech signals may provide clues to their aversion to the highly intonated speech signals typical of motherese.

Recent studies extend the findings on children with autism to word processing using ERP measures (Coffey-Corina, Padden, Kuhl, & Dawson, 2008). In this study, 24 toddlers with autism spectrum disorders between 18 and 31 months of age were separated into high-functioning and low-functioning subgroups defined by the severity of their social symptoms. ERP measures were recorded in response to known words, unknown words, and words played backward. They were compared to ERPs elicited from a group of 20 typically developing toddlers between the ages of 20 and 31 months of age.

The results for typically developing toddlers showed a highly localized response to the difference between known and unknown words at a left temporal electrode site (T3) in the 200–500-ms and 500–700-ms windows (figure 57.5*A*). These data replicate previous data on typically developing children published by Mills and colleagues (1993) and indicate that highly focalized responses are a marker of increasing developmental sophistication in the processing of words in typically developing children. It was therefore of interest to observe that toddlers with ASD showed a very diffuse response to known and unknown words. Known words showed a greater negativity than unknown words across all electrode sites, and at a later latency than age-matched typically developing children (figure 57.5*B*). Both the more diffuse pattern of brain activation and responses with longer latencies are patterns observed in younger typically developing children (Mills et al., 1997).

Replicating the pattern seen in the studies of phonetic perception in children with autism, the word-processing results for children with ASD differed markedly depending on the children's social skills. High-functioning toddlers with ASD produced ERP responses that were similar to those of typically developing children—they exhibited a localized left-hemisphere response to known and unknown words. Significant word-type effects were observed only at the left parietal electrode site (P3) in the 200–500-ms time window (figure 57.5*C*). In contrast, ERP waveforms of low-functioning toddlers with ASD exhibited a diffuse response to words. Known words were significantly more negative than unknown words at multiple electrode sites and in all measurement windows (figure 57.5*D*).

The idea that ERP measures in response to syllables and words may allow us to predict future language outcomes in young children with ASD is exciting. Toward that end, we note that children with ASD exhibited highly significant correlations between their ERP components at the initial test time and their verbal IQ scores measured one year after ERP data collection (figure 57.6).

In new studies with the siblings of children with autism spectrum disorder, we are now exploring whether these early brain and behavioral responses to syllables, and listening preferences for speech, are diagnostic markers for autism. The interest in these measures is that they can be used reliably in infants as early as 6 months of age, an age at which intervention measures might be more effective in changing the course of development for children at risk for autism.

Mirror neurons and shared brain systems

Neuroscience studies that focus on shared neural systems for perception and action have a long tradition in speech research (Fowler, 2006; Liberman & Mattingly, 1985). The discovery of *mirror neurons* for social cognition (Gallese 2003; Meltzoff & Decety, 2003; Pulvermuller, 2005; Rizzolatti, 2005; Rizzolatti & Craighero, 2004) has reinvigorated this tradition. Neuroscience studies using speech and whole-brain imaging techniques have the capacity to examine the origins of shared brain systems in infants from birth (Bosseler et al., 2008; Imada et al., 2006).

In speech, the theoretical linkage between perception and action came in the form of the original *motor theory* (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) and in a different formulation of the direct perception of gestures, named *direct realism* (Fowler, 1986). Both posited close interaction between speech perception and speech production. The perception-action link for speech was viewed as innate by the original motor theorists (Liberman & Mattingly, 1985). Alternatively, it was viewed as forged early in development through experience with speech motor movements and their auditory consequences (Kuhl & Meltzoff, 1982, 1996). Two new infant studies have shed some light on the developmental issue.

Imada and colleagues (2006) used magnetoencephalography (MEG), studying newborns, 6-month-old infants, and 12-month-old infants while they listened to nonspeech signals, harmonics, and syllables (figure 57.7). Dehaene-Lambertz and colleagues (2006) used fMRI to scan 3-month-old infants while they listened to sentences. Both studies show activation in brain areas responsible for speech production (the inferior frontal, Broca's area) in response to auditorally presented speech. Imada and colleagues reported synchronized activation in response to speech in auditory and motor areas at 6 and 12 months, and Dehaene-Lambertz and colleagues reported activation in motor speech areas in response to sentences in 3-month-olds.

Is activation of Broca's area to the pure perception of speech present at birth? Newborns tested by Imada and colleagues showed no activation in motor speech areas for any signals, whereas auditory areas responded robustly to all signals, suggesting the possibility that perception-action

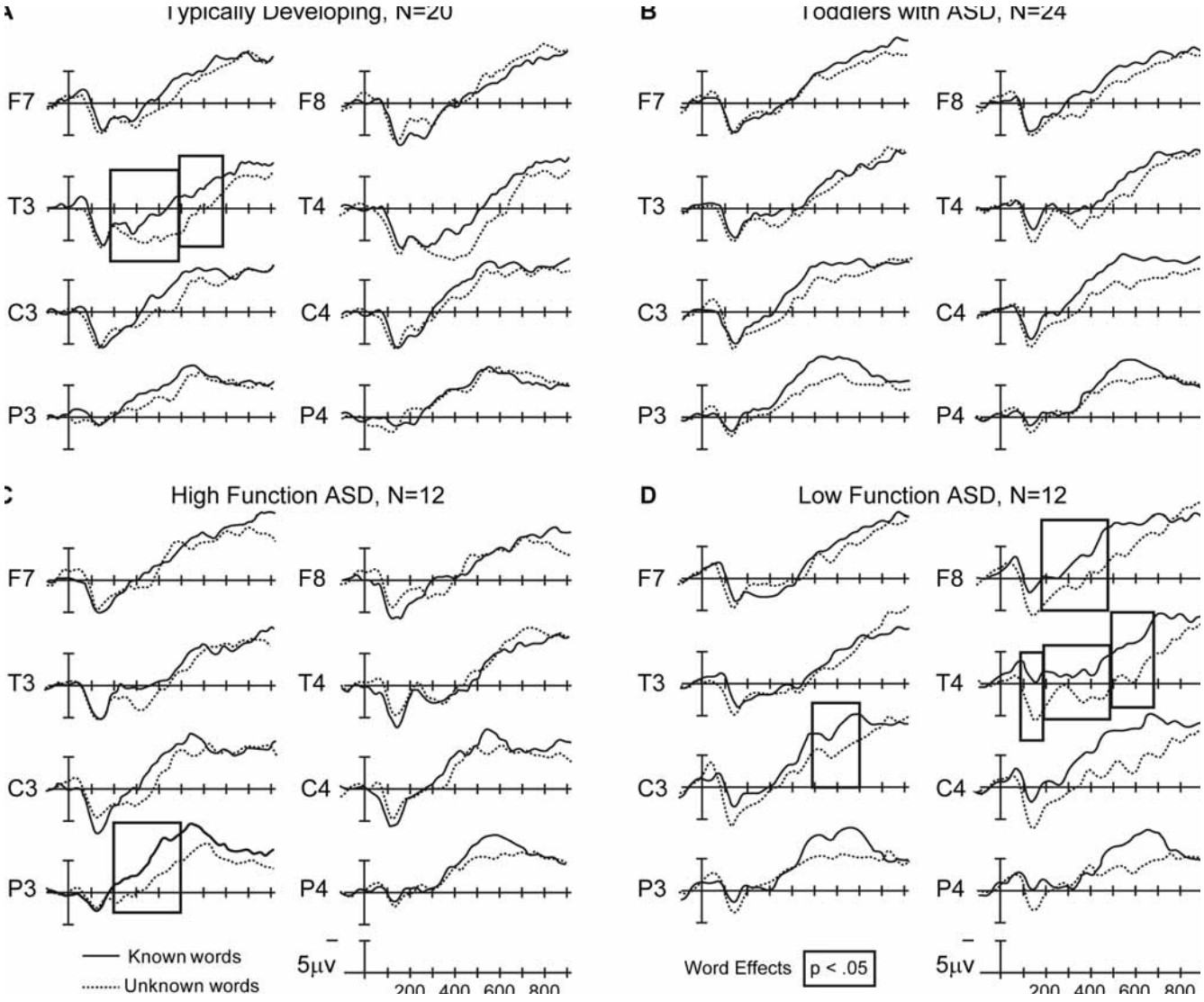


FIGURE 57.5 Group data showing ERP waveforms for (*A*) typically developing toddlers and (*B*) toddlers with autism spectrum disorder. TD toddlers exhibit a localized response with significant differences between known and unknown words at the left temporal electrode site (T3). Toddlers with ASD exhibit a diffuse response to known and unknown words, but the differences are significant across all electrode sites in the 500–700-ms measurement window. (*C*) Subgroup analysis shows that ERP waveforms for high-functioning toddlers with ASD exhibit a localized response with significant differences between known and unknown words at a parietal electrode site in the left hemisphere (P3), similar to typically developing children. (*D*) Low-functioning toddlers with ASD exhibit a diffuse response to known and unknown words with significant differences in multiple time windows and electrode sites, and a significant effect when collapsed across all electrode sites in the 500–700-ms measurement window. (From Coffey-Corina, Padden, Kuhl, & Dawson, 2008.)

linkages for speech develop by 3 months of age as infants produce vowellike sounds. But further work must be done to answer the question. How the binding of perception and action takes place, and whether it requires experience, is one of the exciting questions that can now be addressed with infants from birth using the tools of modern neuroscience.

We now know a great deal about the linkages and the circuitry underlying language processing in adults (Kuhl & Damasio, in press). What is unknown, but waiting to be discovered, is the state of this circuitry at birth and how refined connections are forged in early infancy as perception and action are jointly experienced.

Bilingual infants: two languages, one brain

One of the most interesting questions is how infants map two distinct languages in the brain. From phonemes to words, and then to sentences, how do infants simultaneously bathed in two languages develop the neural networks necessary to respond in a nativelike manner to two different codes?

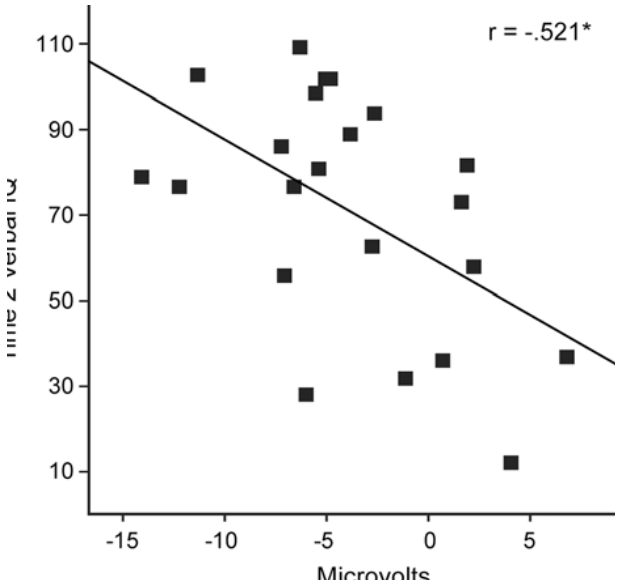


FIGURE 57.6 Predictive correlations for children with ASD between the mean amplitude of ERPs to known words at the left parietal electrode site (P3) and verbal IQ measured one year later. A more negative response predicted significantly higher verbal IQ ($r = -.521$, $p = .013$). (From Coffey-Corina, Padden, Kuhl, & Dawson, 2008.)

Bilingual language experience could potentially have an impact on development—because the learning process requires the development of two codes and because it could take a longer period of time for sufficient data from both languages to be experienced than in the monolingual case. Infants learning two first languages simultaneously might reach the developmental change in perception at a later point in development than infants learning either language monolingually. This difference could depend on such factors as the number of people in the infants' environment producing the two languages in speech directed toward the child and the amount of input they provide. These factors could change the rate of development in bilingual infants.

There are very few studies that address this question thus far, and the data that do exist provide somewhat mixed results. Some studies suggest that infants exposed to two languages show later acquisition of phonetic skills in the two languages when compared to monolingual infants (Bosch & Sebastian-Galles, 2003a, 2003b). This is especially the case when infants are tested on contrasts that are phonemic in only one of the two languages; this has been shown both for vowels (Bosch & Sebastian-Galles, 2003b) and consonants (Bosch & Sebastian-Galles, 2003a). However, other studies report no change in the timing of the developmental transition in phonetic skills in the two languages of bilingual infants (Burns, Yoshida, Hill, & Werker, 2007; Sundara, Polka, & Molnar, 2008). For example, Sundara and colleagues, testing monolingual English and monolingual French as well as bilingual French-English infants, examined discrimination of dental (French) and alveolar (English) consonants. They demonstrated that at 6–8 months, infants in all three language groups succeeded; at 10–12 months, monolingual English infants and French-English bilingual infants, but not monolingual French infants, distinguished the English contrast. Thus bilingual infants performed on par with their English monolingual peers and better than their French monolingual peers. Moreover, data from an ERP study of Spanish-English bilingual infants show that, at both 6–9 and 9–12 months of age, bilingual infants show MMN responses to both Spanish and English phonetic contrasts (Rivera-Gaxiola & Romo, 2006), distinguishing them from English-learning monolingual infants who fail to respond to the Spanish contrast at the later age (Rivera-Gaxiola, Silva-Pereyra, et al., 2005).

ERP studies on word development in bilingual children have just begun to appear. Conboy and Mills (2006) recorded ERPs to known and unknown English and Spanish words in bilingual children at 19–22 months. Expressive vocabulary sizes were obtained in both English and Spanish, and were used to determine language dominance for each child. A conceptual vocabulary score was calculated by summing the total number of words in both languages and then subtracting the number of times a pair of conceptually equivalent words (e.g., "*water*" and "*agua*") occurred in the two languages.

ERP differences to known and unknown words in the dominant language occurred as early as 200–400 and 400–600 ms in these 19–22-month-old infants, and were broadly distributed over the left and right hemispheres, resembling patterns observed in younger (13–17-month-old) monolingual children (Mills et al., 1997). In the nondominant language of the same children, these differences were not apparent until late in the waveform, from 600 to 900 ms. Moreover, children with high versus low conceptual vocabulary scores produced greater responses to known words in the left hemisphere, particularly for the dominant language (Conboy & Mills, 2006).

Research has just begun to explore the nature of the bilingual brain, and it is one of the areas in which neuroscience techniques will be of strong interest. Using whole-brain imaging, we may be able to understand whether learning a second language at different ages—in infancy as opposed to adulthood—recruits different brain structures. These kinds of data may play a role in our eventual understanding of the "critical period" for language learning.

Conclusions

Knowledge of infant language acquisition is now beginning to reap benefits from information obtained by experiments that directly examine the human brain's response to

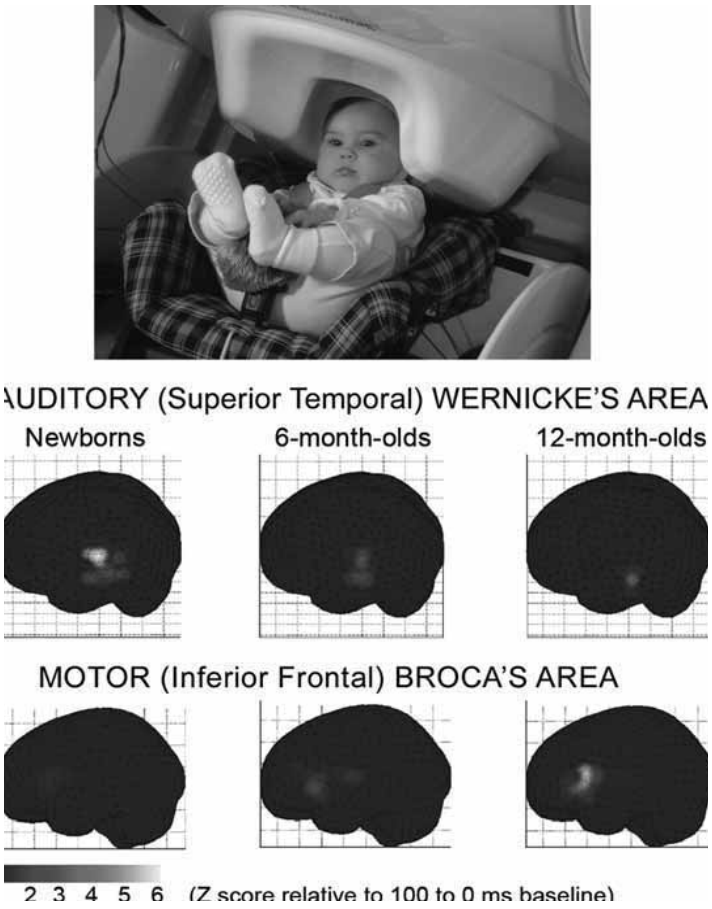


FIGURE 57.7 Neuromagnetic signals were recorded using MEG in newborns, 6-month-old infants, and 12-month-old infants while listening to speech (shown) and nonspeech auditory signals. Brain activation recorded in auditory (*top row*) and motor (*bottom row*) brain regions revealed no activation in the motor speech areas in the newborn in response to auditory syllables. However, activation increased in the motor areas in response to speech (but not nonspeech) in 6- and 12-month-old infants that was temporally synchronized between the auditory and motor brain regions. (From Imada et al., 2006.) (See color plate 74.)

linguistic material as a function of experience. EEG, MEG, fMRI, and NIRS technologies—all safe, noninvasive, and proven feasible—are now being used in studies with very young infants, including newborns, as they listen to the phonetic units, words, and sentences of a specific language. Brain measures now document the neural signatures of learning as early as 7 months for native-language phonemes, 9 months for familiar words, and 30 months for semantic and syntactic anomalies in sentences. Studies show continuity from the earliest phases of language learning in infancy to the complex processing evidenced at the age of three when all typically developing children show the ability to carry on a sophisticated conversation. Individual variation in language-specific processing at the phonetic level—at the cusp of the transition from Phase 1, in which all phonetic contrasts are discriminated, to Phase 2, in which infants focus on the distinctions relevant to their native language—is strongly linked to infants' abilities to process words and sentences two years later. This finding is important theoretically but is also vital to the eventual use of these early speech precursors to diagnose children with developmental disabilities that involve language. In fact, new studies suggest the possibility that early measures of the brain's responses to speech may provide a diagnostic marker for autism spectrum disorder. The fact that language experience affects brain processing of both the signals being learned (native patterns) and the signals to which the infant is not exposed (nonnative patterns) may play a role in our understanding of the brain mechanisms underlying the critical period. At the phonetic level, the data suggest that learning itself, not merely time, may contribute to the critical-period phenomenon. Whole-brain imaging now allows us to examine multiple brain areas during speech processing, including both auditory and motor brain regions, revealing the possible existence of a shared brain system (a "mirror" system) for speech. Research has also begun to use these measures to understand how the bilingual brain maps two distinct languages. Answers to the classic questions about the unique human capacity to acquire language will be enriched by studies that utilize the tools of modern neuroscience to peer into the infant brain.

ACKNOWLEDGMENTS The author is supported by the National Science Foundation's Science of Learning Center grant to the University of Washington LIFE Center (SBE 0354453), by grants from the National Institutes of Health (HD 37954; MH066399; HD34565; HD55782), by core grants (P30 HD02274; P30 DC04661), and by a grant from the Cure Autism Now Foundation. This chapter updates the information in Kuhl, Conboy, Padden, Riveva-Gaxiola, and Nelson, *Philosophical Transactions of the Royal Society of London [Biology]* (2008) and Kuhl and Rivera-Gaxiola, *Annual Review of Neuroscience* (2008).

REFERENCES

Aslin, R. N., & Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants: Promise, prospects, and challenges. *J. Biomed. Opt., 10*, 011009.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cogn. Dev., 17*, 1265–1282.

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *J. Child Lang., 32*, 159–173.

Best, C. C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways, *Lang. Speech, 46*, 183–216.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Dev., 70*, 636–644.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychol. Sci., 16*, 298–304.

Bortfeld, H., Wruck, E., & Boas, D. A. (2007). Assessing infants' cortical response to speech using near-infrared spectroscopy. *NeuroImage, 34*, 407–415.

Bosch, L., & Sebastian-Galles, N. (2003a). *Language experience and the perception of a voicing contrast in fricatives: Infant and adult data.* Paper presented at the Proceedings of the International Congress of Phonological Sciences, Barcelona.

Bosch, L., & Sebastian-Galles, N. (2003b). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Lang. Speech, 46*, 217–243.

Bosseler, A. N., Imada, T., Pihko, E., Mäkelä, J., Taulu, S., Ahonen, A., et al. (2008). Neural correlates of speech and non-speech processing: Role of language experience in brain activation. *J. Acoust. Soc. Am., 123*, 3333.

Brainard, M. S., & Knudsen, E. I. (1998). Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *J. Neurosci., 18*, 3929–3942.

Brooks, R., & Meltzoff, A. N. (2008). Gaze following and pointing predicts accelerated vocabulary growth through two years of age: A longitudinal growth curve modeling study. *J. Child Lang., 35*, 207–220.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*, 1435.

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Appl. Psycholinguist., 28*, 455–474.

Canseco-Gonzalez, E. (2000). Using the recording of event-related brain potentials in the study of sentence processing. In *Language and the brain: Representation and processing* (pp. 229–266). San Diego: Academic Press.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Dev. Sci., 11*, 282–298.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nat. Neurosci., 1*, 351–353.

Cheour, M., Imada, T., Taulu, S., Ahonen, A., Salonen, J., & Kuhl, P. (2004). Magnetoencephalography is feasible for infant assessment of auditory discrimination. *Exp. Neurol., 190*, 44–51.

Chomsky, N. (1959). Review of *Verbal behavior* by B. F. Skinner. *Language, 35*, 26–58.

Coffey-Corina, S., Padden, D., Kuhl, P. K., & Dawson, G. (2007). *Electrophysiological processing of single words in toddlers and school-age children with autism spectrum disorder.* Paper presented at the Annual Meeting of the Cognitive Neuroscience Society, New York.

Coffey-Corina, S., Padden, D., Kuhl, P. K., & Dawson, G. (2008). ERPs to words correlate with behavioral measures in children with autism spectrum disorder. *J. Acoust. Soc. Am., 123*, 3742.

Conboy, B. T., Brooks, R., Taylor, M., Meltzoff, A. N., & Kuhl, P. K. (2008). *Joint engagement with language tutors predicts brain and behavioral responses to second-language phonetic stimuli.* Paper presented at the XVIth Biennial International Conference on Infant Studies, Vancouver, BC.

Conboy, B. T., & Kuhl, P. K. (2007). *ERP mismatch negativity effects in 11-month-old infants after exposure to Spanish.* Paper presented at the Society for Research in Child Development, Boston.

Conboy, B. T., & Mills, D. L. (2006). Two languages, one developing brain: Event-related potentials to words in bilingual toddlers. *Dev. Sci., 9*, F1–F12.

Conboy, B. T., Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2008). Event-related potential studies of early language processing at the phoneme, word, and sentence levels. In A. D. Friederici & G. Thierry (Eds.), *Early language development*, Vol. 5: *Bridging brain and behavior: Trends in language acquisition research* (pp. 23–64). Amsterdam: John Benjamins.

Conboy, B. T., Sommerville, J., & Kuhl, P. K. (2008). Cognitive control skills and speech perception after short-term second language experience during infancy. *J. Acoust. Soc. Am., 123*, 3581.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform., 14*, 113–121.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science, 298*, 2013–2015.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., & Roche, A. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc. Natl. Acad. Sci. USA, 103*, 14240–14245.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303–306.

Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First Lang., 25*, 219–234.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Lang. Speech, 46*, 245–264.

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., et al. (1993). *MacArthur communicative development inventories: User's guide and technical manual.* San Diego: Singular Publishing Group.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Dev. Psychol., 42*, 98–116.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *J. Phonetics, 14*, 3–28.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Percept. Psychophys., 68*, 161–177.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci., 6*, 78–84.

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: From syllables to sentences. *Trends Cogn. Sci., 9*, 481–488.

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cereb. Cortex, 16*, 1709–1717.

Friedrich, M., & Friederici, A. D. (2005). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *NeuroReport, 16*, 653–656.

Friedrich, M., & Friederici, A. D. (2006). Early N400 development and later language acquisition. *Psychophysiology, 43*, 1–12.

Gallese, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 358*, 517–528.

Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Dev. Psychol., 40*, 621–632.

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annu. Rev. Psychol., 54*, 91–114.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *J. Cogn. Neurosci., 16*, 1302–1318.

Hahne, A., & Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cogn. Brain Res., 11*, 199–212.

Halle, P. A., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behav. Dev., 17*, 119–129.

Harris, A. M. (2001). Processing semantic and grammatical information in auditory sentences: Electrophysiological evidence from children and adults. *Diss. Abstr. Int., 61*, 6729B.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behav. Dev.* [Epub ahead of print. doi:10.1016/j.infbeh.2009.03.004.]

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Dev. Neuropsychol., 8*, 203–241.

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neurosci. Res., 54*, 276–280.

Imada, T., Bosseler, A. N., Taulu, S., Pihko, E., Mäkelä, J., Ahonen, A., et al. (2008). Magnetoencephalography as a tool to study speech perception in awake infants. *J. Acoust. Soc. Am., 123*, 3742.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *NeuroReport, 17*, 957–962.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *J. Mem. Lang., 44*, 548–567.

Jusczyk, P. W. (1997). Finding and remembering words: Some beginnings by English-learning infants. *Durr. Dir. Psychol. Sci., 6*, 170–174.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cogn. Psychol., 29*, 1–23.

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science, 277*, 1984–1986.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nat. Rev. Neurosci., 5*, 831–841.

Kuhl, P. K. (2007). Is speech learning "gated" by the social brain? *Dev. Sci., 10*, 110–120.

Kuhl, P. K. (2008). Linking infant speech perception to language acquisition: Phonetic learning predicts language growth. In P. McCardle, J. Colombo, & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research directions* (pp. 213–243). New York: Lawrence Erlbaum.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*, 684–686.

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. *Dev. Sci., 8*, F1–F12.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period." *Lang. Learn. Dev., 1*, 237–264.

Kuhl, P. K., Conboy, B. T., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci., 363*, 979–1000.

Kuhl, P. K., & Damasio, A. (in press). Language. In E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. Siegelbaum, & J. Hudspeth (Eds.), *Principles of neural science* (5th ed.). New York: McGraw Hill.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1141.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. Acoust. Soc. Am., 100*, 2425–2438.

Kuhl, P. K., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annu. Rev. Neurosci., 31*, 511–534.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci., 9*, F13–F21.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. USA, 100*, 9096–9101.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., & Connolly, J. F. (2004). Activation in the anterior left auditory cortex associated with phonological analysis of speech input: Localization of the phonological mismatch negativity response with MEG. *Cogn. Brain Res., 21*, 106–113.

Kutas, M. (1997). Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology, 34*, 383–398.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev., 74*, 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1–36.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Dev. Sci., 6*, F1–F10.

Liu, H. -M., Tsao, F.-M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Dev. Psychol. 43*, 912–917.

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. (1995). Infants' recognition of the sound patterns of their own names. *Psychol. Sci., 6*, 314–317.

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 358*, 491–500.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *J. Cogn. Neurosci., 5*, 317–334.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13–20 months. *Dev. Neuropsychol., 13*, 233–237.

Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cogn. Dev., 20*, 19–31.

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *J. Cogn. Neurosci., 16*, 1452–1464.

Molfese, D. L. (1990). Auditory evoked responses recorded from 16-month-old human infants to words they did and did not know. *Brain Lang., 38*, 345–363.

Molfese, D. L., Morse, P. A., & Peters, C. J. (1990). Auditory evoked responses to names for different objects: Cross-modal processing as a basis for infant language acquisition. *Dev. Psychol., 26*, 780–795.

Molfese, D. L., Wetzel, W., & Gill, L. A. (1993). Known versus unknown word discriminations in 12-month-old human infants: Electrophysiological correlates. *Dev. Neuropsychol., 9*, 241–258.

Munte, T. F., Matzke, M., & Johanes, S. (1997). Brain activity associated with syntactic incongruencies in words and psuedo-words. *J. Cogn. Neurosci., 9*, 318–329.

Naatanen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *J. Mem. Lang., 54*, 283–299.

Neville, H. J., Coffey, S. A., Holcomb, P. J., & Tallal, P. (1993). The neurobiology of sensory and language processing in language-impaired children. *J. Cogn. Neurosci., 5*, 235–253.

Newman, R., N., Ratner, B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Dev. Psychol., 42*, 643–655.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance. I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol., 48*, 127–162.

Oberecker, R., & Friederici, A. D. (2006). Syntactic event-related potential components in 24-month-olds' sentence comprehension. *Neuroreport, 17*, 1017–1021.

Oberecker, R., Friedrich, M., & Friederici, A. D. (2005). Neural correlates of syntactic processing in two-year-olds. *J. Cogn. Neurosci., 17*, 1667–1678.

Pena, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., et al. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proc. Natl. Acad. Sci. USA, 100*, 11702–11705.

Pulvermuller, F. (2005). *The neuroscience of language: On brain circuits of words and serial order*. Cambridge, UK: Medical Research Council, Cambridge University Press.

Raizada, R. D. S., Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2009). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences. *Cereb. Cortex*. [Epub ahead of print. doi:10.1093/cercor/bhp076.]

Raudenbush, S. W., Bryk, A. S, Cheong, Y. F., & Congdon, R. (2005). *HLM-6: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International.

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., & Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. *NeuroReport, 16*, 495–498.

Rivera-Gaxiola, M., & Romo, H. (2006). *Infant head-start learners: Brain and behavioral measures and family assessments*. Paper presented at the From Synapse to Schoolroom: The Science of Learning, NSF Science of Learning Centers Satellite Symposium, Society for Neuroscience Annual Meeting, Atlanta.

Rivera-Gaxiola, M., Silva-Pereyra, J., Klarman, L., Garcia-Sierra, A., Lara-Ayala, L., Cadena-Salazar, C., et al. (2007). Principal component analyses and scalp distribution of the auditory P150–250 and N250–550 to speech contrasts in Mexican and American infants. *Dev. Neuropsychol., 31*, 363–378.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Dev. Sci., 8*, 162–172.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anat. Embryol., 210*, 419–421.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci., 27*, 169–192.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Curr. Dir. Psychol. Sci., 12*, 110–114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Silva-Pereyra, J., Conboy, B. T., Klarman, L., & Kuhl, P. K. (2007). Grammatical processing without semantics? An event-related brain potential study of preschoolers using Jabberwocky sentences. *J. Cogn. Neurosci., 19*, 1050–1065.

Silva-Pereyra, J., Klarman, L., Lin, L. J.-F., & Kuhl, P. K. (2005). Sentence processing in 30-month-old children: An ERP study. *NeuroReport 16*, 645–648.

Silva-Pereyra, J., Rivera-Gaxiola, M., & Kuhl, P. K. (2005). An event-related brain potential study of sentence comprehension in preschoolers: Semantic and morphosyntactic processing. *Cogn. Brain Res., 23*, 247–258.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381–382.

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-ð/ in monolingual and bilingual acquisition of English. *Cognition*, *100*, 369–388.

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, *108*, 232–242.

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Dev. Sci.*, *8*, 432–443.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147–166.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychol. Sci.*, *13*, 480–484.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cogn. Psychol.*, *54*, 99–132.

Taga, G., & Asakawa, K. (2007). Selectivity and localization of cortical response to auditory and visual stimulation in awake infants aged 2 to 4 months. *NeuroImage*, *36*, 1246–1252.

Thierry, G. C. A., Vihman, M., & Roberts, M. (2003). Familiar words capture the attention of 11-month-olds in less than 250 ms. *NeuroReport*, *14*, 2307–2310.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychol. Sci.*, *10*, 172–175.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Dev.*, *75*, 1067–1084.

Vihman, M. M., Nakai, S., DePaolis, R. A., & Halle, P. (2004). The role of accentual pattern in early lexical representation. *J. Mem. Lang.*, *50*, 336–353.

Virnes, M., Cardillo, G., Kuhl, P. K., & Movellan, J. R. (2008). *LIFE TDLC collaboration: Social robots for learning language*. Poster presented at the Learning in Formal and Informal Environments Collaboration, NSF Science of Learning Center, University of Washington, Seattle.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cogn. Brain Res.*, *18*, 149–161.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Lang. Learn. Dev.*, *1*, 197–234.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*, 1–30.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.*, *7*, 49–63.

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural com-mitment to language-specific auditory patterns. *NeuroImage*, *26*, 703–720.

58 Genetics of Language

FRANCK RAMUS AND SIMON E. FISHER

ABSTRACT It has long been hypothesized that the human faculty to acquire a language is in some way encoded in our genetic program. However, only recently has genetic evidence been available to begin to substantiate the presumed genetic basis of language. Here we review the first data from molecular genetic studies showing association between gene variants and language disorders (specific language impairment, speech sound disorder, developmental dyslexia), we discuss the biological function of these genes, and we further speculate on the more general question of how the human genome builds a brain that can learn a language.

Since the beginning of the cognitive revolution, it has been hypothesized that the human faculty to acquire a language is "innate," that is, part of our species' biological makeup and, therefore, encoded in some way in our genetic program (Chomsky, 1959). Over the years, a wide variety of arguments have been advanced in support of this view: the universality of some properties of human languages (Chomsky, 1957), the "poverty of the stimulus" available for language acquisition (Chomsky, 1965), the spontaneous emergence of languages (Bickerton, 1984; Goldin-Meadow & Mylander, 1998), biological adaptations such as that of the vocal tract (Lenneberg, 1967), the existence of inherited disorders that may specifically affect language (Gopnik & Crago, 1991), the heritability of language abilities and disorders (Stromswold, 2001), the adaptiveness of language as a communication system (Pinker & Bloom, 1990), and the plausibility of a gradual evolution of the language faculty (Jackendoff, 1999) (on the special topic of language evolution, see chapter 59 in this volume by W. Tecumseh Fitch).

Although the evidence gathered in the last decades in favor of a biological basis of language looks convincing to many scientists, until recently genetic evidence has remained relatively indirect, in the sense that it has not addressed the fundamental questions: If there is a genetic basis for language, then what exactly is there in the human genome that is different from other species and that gives us language? How does it build a brain that can learn a human language?

There is no easy way to obtain a direct answer to these fascinating questions. Genetic differences between species are only beginning to be systematically searched, and the many differences that are found are not straightforwardly identifiable as associated with language (Fisher & Marcus, 2006). However, part of the answer will likely come from addressing a related but different question: What human genetic variations are associated with variations in the ability to learn a language? Indeed, most genetic methods rely on detecting correlations between variations in the genotype and variations in the phenotype. The capacity to acquire spoken language is usually treated as a universal characteristic of our species. Nevertheless, like many other traits, the language abilities that are observed in the human population vary along a normal distribution. Cases in the lower end of the distribution ("disorders") are typically the most informative, as they may highlight causal relationships between genes, brain, and cognition that are often not readily apparent in normal development. Indeed, disorders of language acquisition have so far provided almost all the available data on language genetics. Furthermore, developmental language disorders are diverse, affecting different aspects of language, therefore promising to illuminate putative genetic influences on particular components of language (phonology, morphology, syntax, articulation . . .). Accordingly, this chapter reviews the genetic data gathered on the various types of language-related disorders (specific language impairment, speech sound disorder, developmental dyslexia . . .) and reflects on what they teach us about the genetic basis of language.

Evidence for genetic influences on language

Historically, the first hint at a genetic influence on language abilities came from the observation that language-related disorders tend to run in families (Hallgren, 1950; Morley, 1967; Stephenson, 1907; Tallal et al., 2001): when one person has language problems, the risk in first-degree relatives is around 50%, far above the normal population prevalence. Although the inheritance pattern in many families may appear consistent with autosomal dominant transmission (e.g., the transmission of a dominant gene variant carried by a nonsexual chromosome), this observation is not sufficient to prove genetic involvement, as members of a family share not only genes but also a linguistic environment. It is conceivable that parents with a language disorder would constitute a less favorable environment for the acquisition of

FRANCK RAMUS Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS, CNRS, DEC-ENS, Paris, France
SIMON E. FISHER Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

language by their children, so studies of familial clustering inevitably confound genetic and nongenetic (shared environmental) factors.

Twin and adoption studies are the usual method to try and disentangle genetic and environmental factors. In the most classic twin studies, one compares the concordance of a given disorder (i.e., the probability that the disorder, when present in one twin, is present in the other one) between monozygotic (MZ) and dizygotic (DZ) twins.[1] For instance, in a meta-analysis of twin studies by Stromswold (2001), the concordance of spoken language disorders was found to be around 84% for MZ twins and 48% for DZ twins. Both figures are far above the typical prevalence of spoken language disorders (1–3%), and the substantial difference between MZ and DZ twins can largely be attributed to differences in their genetic similarity. Such concordance measures thus allow estimation of heritability, that is, the proportion of phenotypic variance than can be attributed to genetic variance. Although diagnostic criteria and precise definition of disorder have varied from one twin study to the next, Stromswold's review of the published research estimated heritabilities of 70% for spoken language disorders and 64% for written language disorders (dyslexia). These estimates have not been significantly challenged, either by more recent studies or by adoption studies that rely on slightly different assumptions (Felsenfeld & Plomin, 1997).

Beyond the categorical classification of individuals as having a disorder or not, the same approach can be generalized to any quantitative measure of language abilities (e.g., vocabulary, syntactic or morphological abilities . . .). Then the correlation of quantitative scores (rather than the concordance of disorders) between twins can be compared between MZ and DZ twins, revealing again higher correlations for the former than for the latter, hence a significant heritability of these scores. One advantage of this approach is that, since it does not require twins to have a disorder, it opens the possibility of assessing genetic influences on variations in normal language abilities as well as on more pathological variations. It turns out that the heritability of normal language abilities is typically lower than that of disorders, yet it remains significantly above zero (Colledge et al., 2002; Stromswold, 2001).

Furthermore, quantitative genetic analyses also lend themselves to investigations of specific components of language. As an example, in a recent study including twin pairs with or without language disorders, the heritability of deficits in various language tests varied depending on whether they tapped primarily phonological short-term memory (61%), morphology (74%), syntax (82%), or vocabulary (1%) (Bishop, Adams, & Norbury, 2006). It is also possible to analyze to what extent the covariance between two phenotypic variables is itself due to genetic and to nongenetic variance. It is generally found that most cognitive abilities are correlated and share genetic variance (Oliver & Plomin, 2007). Nevertheless, it is not the case that all cognitive variables share a single genetic source of variance. For instance, in the study by Bishop and colleagues, morphological and syntactic abilities shared a substantial amount of genetic variance (around 40%), but these abilities in turn did not seem to share much genetic variance with phonological abilities. This finding raises the possibility that certain genetic factors might influence differentially the components of language.

There have been huge debates around twin studies and their implications (Gould & Lewontin, 1979; Joseph, 2002). Their assumptions have been questioned, and their heritability estimates have been argued to be inflated. The fact is that there is no "true" value for heritability; this depends on the particular population considered and on the range of genetic and environmental variance that it presents. Nevertheless, absolute heritability estimates do not matter much. Twin and adoption studies have established beyond reasonable doubt that there are significant genetic influences on cognitive performance and on language disorders in particular. The more important matter now is to identify those genetic factors and understand how they exert their effects. The fact that this approach is now bearing fruit provides a post hoc confirmation of heritability.

A series of progressive advances in molecular biology, culminating with the sequencing of the human genome, now make it possible to carry out the appropriate empirical investigations. Several types of approaches can provide relevant data on language genetics, such as the following:

- Linkage studies, carried out on families, typically analyze which chromosomal regions have genetic markers that are inherited more frequently in family members with a language disorder than in those without. The "linked" chromosomal regions may still contain hundreds of genes, many with unknown function, but they help restrict the search space for association studies.
- Association studies look for gene variants that occur more often in affected than in control individuals, usually at the population level. They can lead to identification of an allele of a gene that increases significantly the risk of developing the disorder. In the case of disorders that are common in the population (like specific language impairment, or SLI, and dyslexia), such alleles may be relatively frequent, also appearing in unaffected individuals. These common alleles may have only subtle effects on gene function, such as reducing the amount of a particular protein that is made.
- Occasionally, sequencing of candidate genes in some families can identify rare mutations that co-occur with the disorder and that severely interfere with the function of the gene in question.
- Comparative studies look for a homologous form of a candidate gene in other species. They typically find one

(at least in mammals). They can then analyze the similarity between the sequences in the various species and attempt to reconstitute the evolutionary history of the specific gene variants that have appeared in the human lineage. Moreover, prior knowledge of the gene's function in other species can give the first clues to its role in humans.

• Expression studies investigate the expression pattern of the candidate gene (where and when the protein is synthesized) as another important clue to its function.

• Many other approaches may be used to further investigate the function of a candidate gene: detection of familiar parts in the sequence and comparison with other, similar genes, algorithmic predictions of the shape of the protein, in vitro experiments to study the mechanisms of action of the target protein and its interactions with other molecules, in vivo experiments to study the effects of disrupting its expression, particularly on brain development and function, and so on.

We now turn to the specific results obtained on the different forms of language disorders.

Developmental dyslexia

Developmental dyslexia is by definition a disorder of reading and spelling acquisition, despite adequate intelligence and opportunity, and in the absence of obvious sensory, neurological, or psychiatric disorder. Nevertheless, it has been well established over the last three decades that most cases of dyslexia can be attributed to a subtle disorder of oral language (the "phonological deficit"),[2] whose symptoms happen to surface most prominently in reading acquisition (Lyon, Shaywitz, & Shaywitz, 2003; Ramus, 2003; Snowling, 2000). Therefore dyslexia is expected to ultimately reveal something about genetic factors implicated in language, in particular in phonology. However, both the exact nature of the phonological deficit and its underlying cognitive/neural causes remain unclear.

Indeed, the main symptoms of the "phonological deficit in dyslexia" are poor phonological awareness (the ability to pay attention to and explicitly manipulate speech sounds), poor verbal short-term memory, and slow lexical retrieval (evidenced in rapid naming tasks where subjects must name series of objects, colors, or digits in quick succession). This diversity of impairments has led many researchers to hypothesize that dyslexics' phonological representations are somewhat degraded, fuzzy, or noisy, lacking either in temporal or spectral resolution, or insufficiently attuned to the categories of the native language. This degradation is assumed either to be specific to the speech-processing system (Adlard & Hazan, 1998; Serniclaes, Van Heghe, Mousty, Carré, & Sprenger-Charolles, 2004; Snowling, 2000) or to follow from a lower-level auditory deficit (Goswami et al., 2002; Tallal, 1980). The latter view has been much challenged in recent years (Ramus, 2003; S. Rosen, 2003; S. White, Frith, et al., 2006; S. White, Milne, et al., 2006). As will become apparent later, the neurobiological and genetic data are consistent with the view that an auditory disorder is not necessary to engender a phonological deficit in people with dyslexia (Ramus, 2004). An alternative view is that phonological representations in dyslexia are intrinsically normal and that the observed difficulties in certain (but not all) phonological tasks arise from a deficit in the access to these representations that is particularly recruited for short-term memory and conscious manipulations (Marshall, Harcourt-Brown, Ramus, & van der Lely, submitted; Ramus & Szenkovits, 2008; Szenkovits, Darma, Darcy, & Ramus, submitted). The elucidation of the precise nature of the phonological deficit will therefore determine whether dyslexia can inform us on the links between genes and phonology per se, or rather between genes and some cognitive processes operating on phonological representations.

In the 1970s, Galaburda and colleagues began to dissect human brains whose medical records indicated a diagnosis of developmental dyslexia (Galaburda & Kemper, 1979). After dissecting four consecutive brains and finding evidence for abnormalities of neuronal migration in all four, they hypothesized that this was unlikely to occur by chance and that such brain development aberrations might provide an explanation of dyslexia (Galaburda, Sherman, Rosen, Aboitiz, & Geschwind, 1985). Most interestingly, neuronal migration disruptions were found predominantly in left perisylvian areas traditionally associated with language. More specifically, these areas are the left inferior frontal, posterior superior temporal, and supramarginal and angular gyri. Galaburda and colleagues subsequently confirmed these findings in three more brains (Humphreys, Kaufmann, & Galaburda, 1990), as well as the rarity of such abnormalities in control brains (Kaufmann & Galaburda, 1989). Unfortunately, no attempt at an independent replication was ever published, so the dyslexia research community came to consider these findings as intriguing but inconclusive. Nevertheless, brain-imaging studies have largely confirmed structural and functional abnormalities in dyslexics' left perisylvian areas, although at a different level of description. Findings from MRI studies typically consist of reduced gray matter density, reduced anisotropy of the underlying white matter, and hypo- or hyperactivations (Démonet, Taylor, & Chaix, 2004; Eckert, 2004; Temple, 2002). At the moment it is impossible to establish their relationship with putative perturbations of neuronal migration, which are not visible in MRI scans. Quite strikingly, new results emerging from genetic studies suggest a reappraisal of the old neuronal migration hypothesis.

Until recently, linkage studies had provided at least six reliable chromosomal loci suspected to harbor genes

associated with dyslexia, on chromosomes 1, 2, 3, 6, 15, and 18 (Fisher & DeFries, 2002; Grigorenko, 2003). Now six genes showing association with dyslexia have been identified in some of these loci: DYX1C1 on 15q21 (Taipale et al., 2003), KIAA0319 on 6p22 (Cope et al., 2005; Paracchini et al., 2006), DCDC2 a nearby gene also on 6p22 (Meng et al., 2005), ROBO1 on 3p12 (Hannula-Jouppi et al., 2005), and MRPL19 and C2ORF3 on 2p12 (Anthoni et al., 2007). The association of variants in KIAA0319 and DCDC2 with dyslexia has been replicated in at least some independent studies (Harold et al., 2006; Schumacher et al., 2005).

For two of these genes (DYX1C1 and ROBO1), mutations, chromosomal rearrangements, or at least rare patterns of alleles (haplotypes) have been found in the dyslexic members of some isolated families, but these changes are too rare to play a significant role in explaining dyslexia in general. As yet, there is little evidence that more common variants of these genes modulate the susceptibility to dyslexia in the general population (Bellini et al., 2005; Brkanac et al., 2007; Marino et al., 2005; Meng et al., 2005; Scerri et al., 2004; Wigg et al., 2004). As far as the other genes are concerned, the associated variants are alleles that are relatively frequent in the population. Thus the mere possession of such a susceptibility allele is not a necessary and sufficient condition to cause dyslexia. Rather, it increases the probability of developing the disorder. Therefore, as predicted by earlier research (Fisher & DeFries, 2002), it seems that the most common cases of dyslexia belong to the family of "complex genetic diseases" (like diabetes, heart disease, and certain cancers), where multiple genetic factors intervene, interact with each other, and interact with environmental factors, thereby modulating the susceptibility to the disorder. Rather than altering the amino-acid sequence of the protein, such susceptibility alleles typically produce more subtle effects, altering quantitatively the expression of the protein (Hannula-Jouppi et al., 2005; Meng et al., 2005) or the way that this is regulated. Follow-up investigations are necessary to pin down the precise functional role of putative risk alleles by studying more directly the structure of the encoded protein and its subdomains (Tapia-Paez, Tammimies, Massinen, Roy, & Kere, 2008; Velayos-Baeza, Toma, da Roza, Paracchini, & Monaco, 2007; Velayos-Baeza, Toma, Paracchini, & Monaco, 2008), as well as its expression patterns across the cortex and at different stages of brain development. It turns out that genes associated with dyslexia are highly (although not exclusively) expressed in the brain, in the cerebral cortex, and particularly so during fetal development (Fisher & Francks, 2006; Meng et al., 2005; Paracchini, Scerri, & Monaco, 2007).

On top of these relatively classic functional studies, LoTurco and colleagues have used a particularly innovative technique to study the role of three of these genes in brain development (Bai et al., 2003). They have produced "functional knockout" rats using in vivo RNA interference. This technique allowed them to specifically block the translation of the gene of interest, in vivo, locally, and at a chosen stage of development (indeed, in utero during neuronal migration). Using this technique, they showed that DYX1C1 is involved in radial neuronal migration and that the part of the protein that is truncated in a Finnish dyslexic family (Taipale et al., 2003) is necessary and sufficient for normal neuronal migration (Wang et al., 2006). They have further shown that cortical ectopias (like the ones observed in dyslexic brains) sometimes occur as a result of the DYX1C1-induced disruption of neuronal migration, and that more generally the laminar organization is locally disrupted, with a distribution of neurons skewed in favor of layers I and II as well as toward the white matter (G. Rosen et al., 2007). The same team has been able to conduct similar studies on both DCDC2 (Burbridge et al., 2008; Meng et al., 2005) and KIAA0319 (Paracchini et al., 2006), again concluding that these genes are likely to be crucial for neuronal migration and the laminar organization of the cortex. Finally, ROBO1 is a homologue of a well-known drosophila gene that is involved in interhemispheric axon guidance and in the migration of cortical interneurons (Andrews et al., 2008; Lopez-Bendito et al., 2007).

A gene will often play multiple roles depending on cellular/developmental context and can be involved in many different processes, but it is striking that functional links to neuronal migration have been uncovered for each of the candidate genes we have described. It would seem a priori highly unlikely that the first four genes associated with developmental dyslexia should all be implicated in this particular aspect of neurodevelopment. The fact that they are suggests that there is indeed a real link between disturbances of neuronal migration and dyslexia, at least in a significant proportion of cases. Thus, 20 years after the first postmortem studies, the emerging genetic findings are remarkably consistent with the original hypothesis of Galaburda and colleagues (Ramus, 2006a), suggesting a relatively coherent account of the etiology of dyslexia, which can be summarized as follows. Certain variants (alleles or mutations) of particular genes increase the susceptibility to disruptions of neuronal migration, sometimes engendering ectopias or microgyri but, most importantly, locally disrupting the laminar organization of the cortex. Through mechanisms that are not yet understood, these disruptions may, in certain individuals, accumulate in left perisylvian areas that are involved in speech processing and phonology, and that are later recruited for reading acquisition. The disruption of these areas also surfaces more macroscopically in the MRI in the form of reduced gray matter density and reduced anisotropy of the underlying white matter. It engenders subtle deficits of phonological abilities that may have little consequence on the acquisition of oral language, but mani-

fest most remarkably during the acquisition of written language, which recruits those abilities particularly intensively (Galaburda, LoTurco, Ramus, Fitch, & Rosen, 2006; Ramus, 2004). There may be alternative neurogenetic pathways that lead to dyslexia and that remain to be uncovered. However, the convergence of data from multiple lines of investigation makes this neuronal migration model particularly compelling as at least one highly testable account of dyslexia etiology.

Specific language impairment

Specific language impairment (SLI) is a disorder of language acquisition that can be attributed neither to mental retardation, nor to other known pathologies (autism, brain lesion, epilepsy, deafness . . .), nor to environmental deprivation or disadvantage. Children with SLI show heterogeneous profiles, but typically have their language development delayed, with reduced vocabulary, reduced expression and/or comprehension abilities, reduced verbal short-term memory, and persistent production of ungrammatical patterns affecting both syntax (sentence structure) and morphology (e.g., verb inflections, gender, plural or case marking) (Leonard, 1998).

At a cognitive level, the most straightforward hypothesis is that children with SLI have deficits in one or several components of language, including syntax, morphology, phonology, the lexicon, and their interfaces (van der Lely, 2005). The precise combination of deficits in a given child, plus the interaction between different language abilities throughout development, would produce the particular cognitive profile presented by the child. An alternative view is that linguistic deficits arise either from a perceptual (auditory) deficit (Tallal & Gaab, 2006; Tallal & Piercy, 1973) or from a more general cognitive deficit (Leonard, 1998; Tomblin & Pandich, 1999). Again, this debate is quite controversial and goes well beyond the present chapter, so we refer the reader to the appropriate literature (Bishop, Adams, Nation, & Rosen, 2005; Ramus, 2004; S. Rosen, 2003; Tallal, 2004; Tallal & Gaab, 2006; van der Lely, 2005; van der Lely, Rosen, & Adlard, 2004; van der Lely, Rosen, & McClelland, 1998). For the purpose of the present discussion, while leaving the precise nature of impairments open, we assume that deficits can have differential impacts on aspects of language. As we will see, this view is at least consistent with the available neurobiological and genetic data.

The overall picture provided by neurobiological data, although far from being clear and consistent, is that loosely defined language-related brain areas are disrupted or differently organized in children with SLI. The most frequent MRI findings have concerned asymmetries between left and right perisylvian areas. The inferior frontal gyrus (IFG: Broca's area) and the planum temporale, generally found to be larger on the left than on the right, show a reduced or reversed asymmetry in people with SLI (De Fossé et al., 2004; Gauger, Lombardino, & Leonard, 1997; Plante, Swisher, Vance, & Rapcsak, 1991). An extra sulcus in the left IFG has also been reported in some individuals with SLI (Clark & Plante, 1998). In addition, it has been suggested that children with SLI present a broader pattern of deviant asymmetries, again in favor of the right hemisphere on average (Herbert et al., 2005). Affected children have also been shown to have a larger total brain volume, as a result of a substantial increase in white matter volume, while the cerebral cortex and the caudate nucleus are relatively smaller (Herbert et al., 2003). Finally, it should be noted that in Galaburda's dissection studies, three to four of the seven patients showed, on top of dyslexia, some form of language delay or disorder (Galaburda et al., 1985; Humphreys et al., 1990). Therefore it is not impossible that the same set of neuronal migration disruptions, perhaps located slightly differently, might lie at the heart of SLI as well as of dyslexia (Ramus, 2004, 2006b). However, there is no direct evidence for that in the case of SLI.

At the genetic level, thus far the search for genes associated with SLI has been less successful than for dyslexia. Nevertheless, there are quite a few interesting results to mention. Familial transmission of language disorders is widely reported, and one study has also reported that atypical perisylvian asymmetry patterns can be found in the relatives of children with SLI (Plante, 1991), suggesting that the transmission of neuroanatomical phenotypes underlies that of behavioral phenotypes. Twin studies also have applications beyond simple heritability estimations. Analyzing correlations between the performance of one twin in a given test and the other twin in a different test allows one to estimate whether the same sources of genetic variance underlie both capacities. One study thus found that syntactic and morphological abilities (typically measured, in English, by the ability to form the past tense of verbs) share some of their genetic variance, but phonological short-term memory and morphological abilities do not (Bishop et al., 2006). This finding suggests that some genetic factors may have differential effects on distinct aspects of language. In a similar vein, another study of children with SLI found that deficits in phonological tests (nonword repetition) are highly heritable, while impairments on a popular auditory processing test do not show significant evidence of genetic influence (Bishop et al., 1999). This finding casts further doubt on the idea that language and phonological deficits necessarily originate from low-level perception.

Finally, genomewide linkage studies of SLI have converged on three main linkage sites: one named SLI1 on chromosome 16, another named SLI2 on chromosome 19 (SLI Consortium, 2002, 2004), and a third one on chromosome 13 (Bartlett et al., 2003, 2002). So far no candidate gene

has been localized in any of these regions, and further mapping studies are under way. However, one recent investigation employed an alternative strategy to traditional mapping, using functional genetic analyses of a monogenic speech and language disorder (described further later on) to identify novel candidates for involvement in SLI. This approach enabled successful identification of the first gene to be significantly associated with language deficits in children with SLI (Vernes et al., 2008). The gene, called CNTNAP2 (located on chromosome 7q35), is strongly down-regulated by the FOXP2 transcription factor in neurons (see the section on developmental verbal dyspraxia) and is a member of the neurexin family, a set of proteins implicated in synaptic adhesion (Dean & Dresbach, 2006). Its association with SLI remains to be replicated. It is worth noting that none of the known SLI linkage sites overlap with those reported for dyslexia, despite frequent comorbidity and similar neurological findings. However, there is notable overlap with autism linkage sites. Furthermore, CNTNAP2 has been associated with autism in several studies (Alarcon et al., 2008; Arking et al., 2008; Bakkaloglu et al., 2008). This issue will be further discussed in the comorbidity subsection.

Speech sound disorder

Although most children make speech errors when they begin to speak, children with speech sound disorder (SSD) present with persistent difficulties in the accurate and intelligible production of speech sounds within words. Their prevalence is estimated to be around 15% of 3-year-old children and 3.8% of 6-year-olds (Shriberg, Tomblin, & McSweeny, 1999). Typically some speech sounds are omitted or mapped to other sounds (this disorder is different from stuttering). The definition of SSD does not commit to a particular locus for the underlying deficit (phonological or articulatory), and it is likely that the population is heterogeneous in this respect. Unfortunately, cognitive studies of SSD are currently insufficient to provide a clear typology and shed more light on the precise nature of the deficits. It should be noted that the field of (normal) child language is itself plagued by the issue of whether deviant speech productions should be attributed to constraints in articulatory skills or to stages of phonological acquisition (Ramus et al., in press).

The brain basis of SSD has to our knowledge not been investigated independently from that of SLI or dyslexia. There have been, however, genetic linkage studies. Investigations have tended to focus on the chromosomal regions implicated in dyslexia and, intriguingly, have thereby uncovered SSD linkages on the dyslexia-related sites of chromosomes 3, 6, and 15 (Stein et al., 2006, 2004). One possible reason for this result is that there is comorbidity between dyslexia and SSD, so that a fair proportion of preschool children who are diagnosed with SSD grow up to become dyslexic. Thus cohorts of children with SSD participating in genetic studies may well be largely composed of dyslexic children. Another more interesting potential explanation is that, beyond actual comorbidity, common biological factors may participate in the etiology of different cognitive deficits. Confirmation of the latter awaits identification of particular allelic variants that play functional roles in both SSD and dyslexia. Curiously, at this point there is less evidence of genetic risk factors that are shared between SSD and SLI, although there may well be functional pathways that are common to both (see the next section).

In conclusion, speech sound disorder has the potential to reveal important information about the genetic bases of phonology and speech articulation. Unfortunately, the findings on SSD in general are rather scarce, so this disorder warrants more investigation. However, one particular form of SSD, namely developmental verbal dyspraxia, is currently at the center of a very fruitful line of research, which is detailed in the next section.

Developmental verbal dyspraxia

Developmental verbal dyspraxia (DVD)—also referred to as childhood apraxia of speech (American Speech-Language-Hearing Association, 2007)—is a speech-sound disorder that leans clearly on the articulation side, involving problems with coordinating and sequencing movements of the tongue, lips, jaw, and palate that cannot be explained by muscle weakness, paralysis, or other overt neurological or physical factors. A diagnosis of DVD can encompass a range of severities and impairments, and there may also be some degree of impairment in performing nonspeech orofacial movements on command, such as puffing out cheeks or licking lips (oral dyspraxia). In recent years substantial advances have been made in understanding one particular genetically mediated subtype of DVD, a rare form of the disorder showing monogenic inheritance (Fisher, Vargha-Khadem, Watkins, Monaco, & Pembrey, 1998; Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001). In this section we focus on the behavioral, cognitive, and neural features of this well-studied subtype, given that its genetic basis has now been firmly established.

Much of our understanding of links between genes and DVD stems from intensive studies of one multigenerational pedigree, known as the KE family, first reported in the early 1990s (Hurst, Baraitser, Auger, Graham, & Norell, 1990). Around half of the members of this family—15 individuals across three successive generations—display a severe speech and language disorder, inherited as a Mendelian trait with an autosomal dominant mode of transmission. While some linguists initially characterized the KE family's disorder as one primarily affecting certain features of grammatical processing (Gopnik, 1990; Gopnik & Crago, 1991), other

researchers noted that the most profound problems were impaired speech articulation reminiscent of DVD (Hurst et al., 1990). Indeed, subsequent reports showed that word and nonword repetition tasks provided the most robust diagnostic marker of the disorder (Vargha-Khadem et al., 1998). Consistent with a diagnosis of DVD, the deficits of affected members are already evident when repeating shorter utterances, but become more dramatic with increases in syllable number and complexity (Watkins, Dronkers, & Vargha-Khadem, 2002). Tests of nonspeech praxis in the KE family indicate reduced performance when making simultaneous and sequential oral movements on command (Alcock, Passingham, Watkins, & Vargha-Khadem, 2000; Vargha-Khadem et al.). This is again reminiscent of other cases of DVD, which (as noted earlier) often show evidence of oral dyspraxia affecting nonspeech movements. Notably, affected members of the KE family are not significantly impaired in making single simple oral movements or in limb praxis, and they do not show gross oromotor dysfunction, for example, in feeding or swallowing (Alcock et al.).

The speech difficulties of the KE family are accompanied by linguistic impairments that are not confined to spoken language or to the expressive domain. For example, affected members perform worse than unaffected members on written tests of verbal fluency and nonword spelling, as well as in lexical decision tasks assessing receptive vocabulary, and they display significant deficits in reception and production of grammar (Watkins, Dronkers, et al., 2002), albeit not as selectively as proposed in initial linguistic studies (Gopnik, 1990). They show difficulties in generating word inflections and derivations, but tests of past-tense production indicate similar levels of deficits for both regular and irregular words, and their receptive impairments extend to syntax at the word-order level (Gopnik & Crago, 1991; Watkins et al.). The relationship between the motoric and linguistic aspects of the disorder in the KE family is the subject of continuing debate. One hypothesis is that a primary deficit in articulation could lead to more general impoverishment in language representation at many other levels (Watkins et al.). However, it is not clear why accurate speech articulation would be necessary to acquire all the other dimensions of language, and indeed it has been shown that it is not (Fourcin, 1975a, 1975b; Lenneberg, 1962; Ramus, Pidgeon, & Frith, 2003). A plausible alternative is that multiple components of language (articulation, phonology, the lexicon, morphology, and syntax) are concurrently affected, without one deficit being responsible for all the others.

The brains of affected people from the KE family appear overtly normal in structure on standard evaluation of MRI scans (Vargha-Khadem et al., 1998). However, statistical comparisons to unaffected members using voxel-based morphometry revealed subtle anomalies affecting multiple brain regions (Belton, Salmond, Watkins, Vargha-Khadem, & Gadian, 2003; Vargha-Khadem et al.; Watkins, Vargha-Khadem, et al., 2002). These include putative abnormalities in cortical language-related regions, with decreased gray matter density in the inferior frontal gyrus (containing Broca's area) and increased density in the posterior portion of the superior temporal gyrus (Wernicke's area). Notably, the sites of pathology suggested by such analyses were not limited to the cerebral cortex, but extended to the cerebellum and the striatum, where there were significant reductions in gray matter density in the caudate nucleus accompanied by increases in the putamen. Functional neuroimaging of the KE family during language tasks identified abnormal patterns of neural activation in the affected members, even under covert (silent) conditions when there was no requirement for spoken output (Liegeois et al., 2003). Broca's area, other cortical language-related regions, and the putamen were significantly underactivated in affected individuals, who showed a more posterior and bilateral pattern of activation than unaffected members of the family. Sites of abnormalities include both areas associated with motor control and areas associated with language, mirroring the co-occurrence of motor and linguistic symptoms at the cognitive level. It has been suggested that abnormalities in development and function of distributed frontostriatal and/or frontocerebellar circuits are responsible for the DVD and accompanying linguistic impairments of the family (Vargha-Khadem, Gadian, Copp, & Mishkin, 2005).

Genomewide scanning of the KE family identified a region of chromosome 7q31 showing highly significant linkage to the disorder (Fisher et al., 1998), which was found to contain at least 70 genes (Lai et al., 2000). The search was cut short by the serendipitous discovery of another child affected with DVD (unrelated to the KE family) who had a gross chromosomal abnormality mapping within the region of interest (Lai et al., 2000, 2001). The child, known as CS, carried a balanced translocation involving exchange of material between chromosomes 5 and 7, with a breakage in the 7q31 band. It was shown that the chromosome 7 breakpoint of this child directly interrupted a novel gene, known as FOXP2 (Lai et al., 2001). Analysis of the gene in the KE family uncovered a heterozygous single-base change in all 15 affected members, which was not found in any unaffected members or in several hundred independent controls (Lai et al., 2001). This mutation was predicted to disrupt the function of the protein encoded by FOXP2, a hypothesis that has since been robustly confirmed (Groszer et al., 2008; Vernes et al., 2006).

FOXP2 encodes a protein belonging to the "Forkhead bOX" (or FOX) family of transcription factors, which act to regulate the expression of suites of genes during embryogenesis and development and in adulthood (Carlsson & Mahlapuu, 2002). The single-base missense mutation in the FOXP2 gene of affected KE family members alters one

amino-acid residue at a crucial part of the DNA-binding domain of the encoded protein (Lai et al., 2001). Functional experiments show that the substitution impedes the DNA-binding ability of the mutated FOXP2 protein, dramatically disturbing its capacity to regulate transcription of downstream targets (Vernes et al., 2006).

Targeted screening of FOXP2 in different disorders has indicated that disruption of this gene is not unique to the KE family and CS case, but still represents only a rare cause of speech and language deficits in the wider population. Initially, comprehensive mutation searches were carried out across all known FOXP2 exons in groups of children with SLI and autism (Newbury et al., 2002; Wassink et al., 2002), syndromes that typically occur in the absence of DVD. These studies concluded that FOXP2 is not a major genetic risk factor for SLI or autism, a finding that has been generally borne out by subsequent work. MacDermot and colleagues (2005) reported the first specific assessment of FOXP2 contribution in a cohort of children diagnosed with DVD. The study screened 49 unrelated probands with a primary diagnosis of DVD and identified three distinct coding changes. One was a heterozygous nonsense mutation predicted to severely truncate the encoded FOXP2 protein, such that it would lack crucial functional domains, including the DNA-binding motif. The nonsense mutation was also found in the proband's affected sister and mother, and was absent from normal controls (MacDermot et al.). Functional analyses suggest that the truncated product is unstable, is mislocalized within the cell, and lacks transcription factor function (Groszer et al., 2008; Vernes et al., 2006). In recent years, cases of gross chromosome abnormalities in which FOXP2 is disrupted or deleted have also been reported, with speech articulation difficulties emerging as a common symptom (Feuk et al., 2006; Shriberg et al., 2006; Zeesman et al., 2006).

Since FOXP2 encodes a transcription factor, functional genomic methods are now being used to successfully identify the downstream target genes that it regulates in neurons (Spiteri et al., 2007; Vernes et al., 2007). Exciting new data from these screening efforts indicate that pathways downstream of this regulatory factor may have broader relevance for language-related disorders, even in the absence of mutations of FOXP2 itself. Vernes and colleagues (2008) identified a novel direct target that is strongly downregulated by FOXP2 in neurons (the CNTNAP2 gene, described earlier) and went on to show that the allelic variants of this target were significantly associated with language impairments in a large cohort of children with typical SLI. Not only do these findings establish a functional genetic link between rare monogenic forms of DVD and common forms of SLI, but similar allelic variants in the target gene are also associated with language deficits in autistic disorder (Alarcon et al., 2008).

FOXP2 is expressed in the brain during embryogenesis and early development, both in humans and in mice (Lai, Gerrelli, Monaco, Fisher, & Copp, 2003). It is not expressed ubiquitously throughout the brain, but localized to a number of structures, including the deep layers of the cerebral cortex, the striatum, the thalamus, the Purkinje cells of the cerebellum, and the inferior olives. Most notably, FOXP2 expression in the caudate nucleus and in the cerebellum coincides with known sites of neuroanatomical anomalies in the KE family. Beyond sensorimotor processing and motor-skill learning, the contribution of these brain regions to language function is becoming more and more appreciated (Booth, Wood, Lu, Houk, & Bitan, 2007; Friederici & Kotz, 2003; Justus, 2004; Marien, Engelborghs, Fabbro, & De Deyn, 2001; Teichmann, Dupoux, Kouider, & Bachoud-Levi, 2006; Ullman, 2001).

More insights into human FOXP2 function have come from animal models. Heterozygous mice carrying the same missense mutation as that found in the human KE family display abnormal synaptic plasticity in neural circuits where FOXP2 is expressed, including loss of long-term depression in parts of the striatum (Groszer et al., 2008). In addition, they show subtle but significant motor-skill learning deficits during species-typical behaviors. Homozygous mouse pups that have no functional FOXP2 have severe motor dysfunction, general developmental delays, and delayed maturation of the cerebellum, and they do not emit innately specified ultrasonic calls on isolation from their mother (Groszer et al.; Shu et al., 2005). They do not survive beyond a month of life. Whether the homozygous mouse phenotype is relevant for understanding the syndrome observed in heterozygous humans remains a controversial question. More convincing evidence of a role for FOXP2 in vocalization skills of nonlinguistic species comes from studies of vocal learning in songbirds (S. A. White, Fisher, Geschwind, Scharff, & Holy, 2006). In particular, zebra finches show changes in FOXP2 expression levels in a key striatal nucleus (called Area X) that appear to correlate with vocal plasticity (Haesler et al., 2004; Teramitsu & White, 2006). Haesler and colleagues (2007) used RNA interference to selectively knock down expression of FOXP2 in Area X of juvenile zebra finches during song learning. This treatment yielded inaccurate and incomplete copying of the tutor's song, which was suggested to show parallels to DVD in humans (Haesler et al., 2007).

Finally, analyses of the evolution of FOXP2 in primates indicated that two amino-acid substitutions occurred on the human lineage after splitting from the chimpanzee, and found evidence of recent Darwinian selection (Enard et al., 2002; Zhang, Webb, & Podlaha, 2002). Although initial studies suggested this accelerated evolution may have occurred within the last 200,000 years of human history (Enard et al.; Zhang et al.), investigations of the gene in bone

samples from Neanderthals indicate that they also carried the human amino-acid substitutions, which would suggest a more ancient origin (at least 300,000–400,000 years) for the changes (Krause et al., 2007). At the moment, nothing is known about the functional consequences of these two amino-acid changes, but this finding raises the possibility that FOXP2 might have acquired new functional roles in humans.

In summary, FOXP2 may simultaneously contribute to human language pathways by at least two routes. The first route is through an evolutionarily conserved role related to motor sequencing and vocal learning, as observed in non-linguistic species (studies of birds and mice). Deficits in these processes are likely to mediate parts of the DVD phenotype associated with FOXP2 disruption. Second, the human version may have putative novel functions that remain to be understood but that might conceivably contribute to more human-specific aspects of language.

Perspectives for language genetics

Comorbidity and Pleiotropy Until now we have largely described the different forms of language disorders as if they were distinct entities; however, this approach is an oversimplification. Many children with SLI, although not all of them, grow up to become dyslexic (Bishop & Snowling, 2004; Flax et al., 2003; Marshall, Harcourt-Brown, Ramus, & van der Lely, in press; McArthur, Hogben, Edwards, Heath, & Mengler, 2000). Some children with dyslexia or SLI also present some form of speech sound disorder, if only in early development (Bishop & Adams, 1990; Shriberg et al., 1999). This pattern of multiple comorbidities is hardly surprising if one considers that the different components of language, albeit functionally independent, may partly depend on each other in the course of development. But beyond this observation, it is likely that comorbidity can be largely ascribed to common underlying biological factors, as is indeed suggested by several lines of converging evidence:

- As we have noted, the neural bases of dyslexia and SLI partly overlap.
- Familial aggregation studies have found that in families having one member with SLI or SSD, the likelihood that other members will show another form of language impairment (whether dyslexia, SLI, or SSD) was increased (Flax et al., 2003; Lewis, 1992).
- Twin studies have provided evidence for shared genetic influences between SSD and dyslexia, suggesting that the cofamiliality has at least partly a genetic basis.
- Genetic linkage sites seem to overlap between dyslexia and SSD. Two caveats, however. First, the fact that linkage sites overlap does not guarantee that a single gene is associated with both disorders: linkage sites may contain many genes, including two affecting different disorders. And indeed none of the genes associated with dyslexia has been associated with SLI or SSD so far. Second, there is no hint as yet of any overlap between dyslexia and SLI linkage sites, a fact that may seem puzzling. However, it is not all that surprising, given the statistical power of most linkage analyses (Marlow et al., 2003), and this gap may well be bridged sooner or later.
- Genetic linkage sites also overlap between SLI and autism. Furthermore, the CNTNAP2 gene, identified as a downstream target of FOXP2, also appears to be associated with common cases of SLI (Vernes et al., 2008), as well as with autistic spectrum disorder (Arking et al., 2008; Bakkaloglu et al., 2008). One study further suggested the association between CNTNAP2 and language abilities in autism, as measured by age at first word (Alarcon et al., 2008). This finding suggests etiological overlaps between SLI and autism.

The possibility that some gene variants might increase the susceptibility to several disorders makes sense in functional terms. For instance, there is no reason to expect that dyslexia is the only disorder arising from slight disturbances in neuronal migration (indeed, others are known, such as nodular periventricular heterotopia). Therefore genes involved in neuronal migration and associated with dyslexia could plausibly be expected to be associated with other disorders such as SLI. Furthermore, genes typically have more than one function, and therefore can have effects on multiple phenotypes: this condition is known as pleiotropy. For instance, all the genes discussed in this chapter are expressed not only in the developing brain, but also in other organs at various stages of life, showing that they have multiple functions, some as remote from cognition as digestion or reproduction.

These considerations have led Kovas and Plomin (2006) to hypothesize that genes affecting cognition are "generalist genes" affecting most cognitive functions and disorders, and indeed that they produce their effects relatively uniformly on a "generalist brain." It is certainly true that many genes affect many brain areas and many cognitive functions, yet the "generalist genes" hypothesis is likely to be an overgeneralization. Some twin studies find that certain cognitive functions share little genetic variance—for instance, phonological and morphosyntactic abilities (Bishop et al., 2006). And although many genes seem to be expressed more or less uniformly across the cortex, few studies have actually compared the expression of the genes of interest across different cortical areas. FOXP2 is a good case in point. It may well have multiple effects on development, but it certainly does not have uniform effects throughout the brain. As we have seen, it is expressed in particular brain areas that turn out to bear a clear relationship with the neurological and

cognitive phenotypes associated with a FOXP2 mutation. This kind of neuroanatomical specificity is not uncommon among transcription factors. Performing a systematic search over more than 1000 known transcription factors, Gray and colleagues (2004) have found 349 whose expression pattern is restricted to specific areas of the mouse brain and which are together sufficient to explain its architecture. Far from being generalist genes, their expression is rather specific and has equally specific functional consequences. Similar considerations hold for CNTNAP2, the only gene so far suggested to be associated with SLI (Vernes et al., 2008), which turns out to demonstrate particularly enriched fetal expression in human frontal cortex (including inferior and middle frontal gyri), as well as in subcortical areas (including the caudate nucleus) (Abrahams et al., 2007).

In the case of genes associated with dyslexia, while expression patterns in human fetal brains are available (Paracchini et al., 2006), comparisons between neocortical areas have been carried out in adult brains only, and with a relatively rough cortical parcellation (lobe by lobe, without distinguishing left from right hemisphere). Yet they do not turn out to be particularly uniform (Meng et al., 2005; Paracchini et al., 2007). Most importantly, the sites of brain disturbance themselves are clearly not uniform, whether one looks at histological studies, brain morphometry, or diffusion tensor imaging. The relationship between genes and neuropathological sites remains to be fully understood. More detailed studies might reveal that genes associated with dyslexia are expressed more in left perisylvian areas, but this possibility can be considered unlikely for genes generally involved in neuronal migration. Then why do the disruptions occur precisely there? One reason could be just chance: in many individuals with the same gene variants, they may by chance occur elsewhere and produce other effects (SLI, SSD, or any other cognitive deficit for that matter). We would see them in left perisylvian areas because we look only at dyslexic individuals. Yet if chance were the only factor at play, one would predict complete cross-transmission between disorders: dyslexic parents would be as likely to beget SLI as dyslexic children. However, this is not the case (Flax et al., 2003; Lewis, 1992). Another possibility would be that left perisylvian areas are, for unrelated (say, vascular) reasons, more vulnerable to all forms of insult, including disturbances of neuronal migration (Geschwind & Galaburda, 1985; McBride & Kemper, 1982). One way or another, neuroanatomical location matters more than anything else for determining the precise nature of a cognitive phenotype.

Another alternative would be that genes implicated in neuronal migration interact with other genes, which do have more specific expression patterns (Ramus, 2004). The combination of certain alleles in these different genes could result in disruptions of neuronal migration confined to certain cortical areas. For instance, a number of genes have been found whose expression is asymmetric between left and right hemispheres in early embryonic development and could thus explain the predominance of certain anomalies on one side or the other. Furthermore, one of these genes (LMO4) is expressed more specifically in perisylvian regions, and more so in the right than in the left hemisphere (Sun et al., 2005). Other genes have been found with expression enriched (or specifically impoverished) in language-relevant areas in midgestation (Abrahams et al., 2007). Alleles of these or similar genes, interacting with alleles of genes associated with neuronal migration, could potentially explain the occurrence of neuronal migration anomalies specifically in left perisylvian regions such as in dyslexia.

In light of the preceding discussion on comorbidity and pleiotropy, one does expect to find genes associated with dyslexia as well as SSD and/or SLI, and perhaps even with other developmental disorders. However, this does not imply that all disorders are the same or that genes are "genes for everything." Not all dyslexic children have SSD or SLI, not all brain areas are involved in all language functions, not all genes have an impact on all brain areas and functions, and therefore it is also to be expected that some genes will be uniquely associated with one disorder, alongside other genes that will be more general susceptibility factors for a certain class of neurodevelopmental disorders.

A "Gene for Language"? When the KE family was first investigated in the early 1990s, speculations about the existence of a "gene for grammar" flourished in the press. The story turned out to be much more complex, and when FOXP2 was discovered more than ten years later, it became clear that it was neither a gene for grammar, nor a gene for language, nor a gene for the brain, nor even a specifically human gene. It is a highly conserved transcription factor, found in similar form in many distantly related vertebrate species, where it is expressed in a range of tissues during embryonic development, postnatally, and in the mature organism, including the lung, heart, and intestines as well as the brain (Bonkowsky & Chien, 2005; Haesler et al., 2004; Lai et al., 2001, 2003). Genes associated with dyslexia and other language disorders are turning out to show similar characteristics. Thus, the very notion of a "gene for something," in particular a gene coding directly, specifically, and uniquely for a given cognitive function, is flawed (Fisher, 2006). But this fact does not mean that the notion of genetic bases of language is itself flawed. Rather it should be understood in less naive ways than it sometimes has been.

The data reviewed in this chapter show that variations in many genes may cause variations in language abilities, and in particular language disorders. Rather than being "genes for language," these genes perform several different functions, in various organs at various stages of development. But they have in common that they have an influence on brain

development and that certain of their variations may alter the development and/or function of particular brain areas, which in turn are useful for some aspects of language acquisition. Thus these genes are necessary for normal language acquisition, but they are of course not sufficient, and furthermore they have not necessarily evolved for the purpose of language acquisition. Some of them (like FOXP2) have indeed undergone some human-specific modifications, apparently under selection pressure, and within a time frame that is compatible with the evolution of language in the human lineage. In such a case it is possible that these changes were one of the steps that made it possible for humans to develop language. Other known genes associated with language disorders also differ slightly between humans and other mammals, but so far there is no evidence that these differences are functionally significant and may have played a role in language evolution (Fisher & Francks, 2006). Nevertheless, this lack of evidence does not make those genes uninteresting.

The language faculty is very unlikely to be an entirely new organ that has appeared from scratch in the human brain (Fisher & Marcus, 2006). Rather, it should be seen as a product of "descent with modification," that is, a new combination of old and possibly new cognitive ingredients (Marcus, 2006). Old ingredients may include auditory perception, primate vocalization, long-term, short-term, and working memory, sequence processing, a conceptual system, and many more. Of course each of these components must have to some extent evolved in human-specific ways in order to be harnessed for linguistic purposes, a fact that implies that some of the genes that were already implicated in the construction of the corresponding brain areas either have undergone some functional changes or have been triggered in new ways by upstream transcription factors and other regulatory elements. Thus even a human gene identical to an ancestral primate version could nowadays be important for language, if for instance it is involved in the construction of a relevant brain area in virtue of being expressed in new ways by a transcription factor such as FOXP2. As for new cognitive ingredients, it is not yet entirely settled what (if anything) should fall into that category. An influential and controversial proposal is that a capacity for recursion is the unique new cognitive ingredient required for language, together with an adaptation of "interfaces" between this new component and the old ones (Fitch, Hauser, & Chomsky, 2005; Hauser, Chomsky, & Fitch, 2002; but see Jackendoff & Pinker, 2005; Pinker & Jackendoff, 2005).

Taking this as a working hypothesis, it is unlikely that such a new cognitive capacity could have evolved overnight thanks to a single mutation. Even if it is truly new in a cognitive sense, it is likely to be much less novel in biological terms. For instance, a change in a single gene producing a signaling molecule (or a receptor, channel, etc.), could lead to creating new connections between two existing brain areas. Even an altogether new brain area could evolve relatively simply by having a modified transcription factor prenatally define new boundaries on the cortex, push around previously existing areas, and create the molecular conditions for a novel form of cortex in Brodmann's sense: still the basic six layers, but with different relative importance, different patterns of internal and external connectivity, and different distributions of types of neurons across the layers. This would essentially be a new quantitative variation within a very general construction plan, requiring little new in terms of genetic material, but this area could nevertheless present novel input/output properties that, together with the adequate input and output connections, might perform an entirely novel information-processing function of great importance to language. Even if the ultimate form of that brain area turns out to require many genetic changes, there is no necessity that all the changes coevolved simultaneously. Once the area is delineated, further genetic changes could progressively shift its boundaries and refine its cellular makeup and thus its information-processing capabilities. Thus even the creation of a new neuroanatomical and cognitive module is not as unlikely as one might imagine and does not require improbable assumptions about dramatic genetic changes. Dramatic effects can be obtained by small changes in the way the construction plan is laid out.

In a nutshell, there is no need of a "gene for language" to explain the genetic basis of language. Having said that, it is now known that some human genes (perhaps 150 to 300) really are human specific, in the sense that they are entirely new concatenations of bits of other genes that have no equivalent in other species (Bailey et al., 2002; Nahon, 2003). Very little is known about those genes, but it is of course possible that one or more of them could have been important in the evolution of the neural bases for language. The point is that even if this is not the case, more standard genetic changes in ancestral genes would still be adequate to explain the emergence of a new cognitive ability such as language.

Perspectives The picture laid out in this chapter is of course very incomplete. Many more genes associated with language disorders remain to be found, and genes associated with normal variations in language abilities remain even to be searched for. Nevertheless, the data that we have discussed are probably a reasonable illustration of what can be expected in the future. We can expect more genes involved in aspects of brain development (neuronal migration being just one possibility), as well as more transcription factors and other genes with a restricted cortical expression that may affect the development of more specific brain areas. Genes involved in neurotransmission, however, are currently out of the picture (although implicated in other disorders such as ADHD), but there is of course no guarantee that they will remain so.

One point that may change is that until now the genetic variations considered have been mostly deletions, insertions, or substitutions of single nucleotides. This approach has led to a pattern where mutations (such as those in FOXP2 or DYX1C1) appear to be scarce, while most of the variation in language abilities seems to be explained by susceptibility alleles that simply modulate the probability of developing the disorder. However, mutation-screening efforts are very preliminary; for instance, the genes already known to be associated with dyslexia have typically not been systematically screened for mutations in most available dyslexia cohorts. Furthermore, a wider range of mutations is now going to be analyzed, such as copy number variants, whereby entire stretches of DNA are sometimes deleted or duplicated, to an extent that previously has been vastly underestimated (Redon et al., 2006; Stranger et al., 2007). Thus there may be etiological mutations in a much higher proportion of individuals with language disorders than has been appreciated before.

One final area where entirely novel results should be expected in the coming years is that of gene-environment interactions. All genetic studies of language disorders have until now focused on detecting main effects of gene variants. This is of course the first step necessary to the identification of candidate genes. However, the effects of genes sometimes differ as a function of other factors, some genetic, some environmental. Evidence for nonadditive effects between genetic and environmental factors has begun to be uncovered in the case of other disorders, such as conduct disorder (Caspi et al., 2002) or depression (Caspi et al., 2003). Does a susceptibility allele for a language disorder produce a different effect depending on the presence of other risk factors (such as mild hearing impairment)? Or on the familial linguistic environment? Or on the language itself? Or on schooling practices? Or symmetrically, does a given environmental factor produce a different effect depending on the genotype of the child? Answers to these fascinating questions are now within arm's reach.

ACKNOWLEDGMENTS FR is supported by Agence Nationale de la Recherche (Genedys) and the European Commission (Neurodys). SEF is a Royal Society Research Fellow, and is also supported by the Wellcome Trust and Autism Speaks.

NOTES

1. Monozygotic MZ twins share 100% of their genome, while dizygotic DZ twins share only 50% of their gene variants (like ordinary siblings). Note that the MZ-DZ twin method usually assumes that environmental factors are not more similar for MZ twins than for DZ twins; this assumption may not necessarily be valid.
2. A minority of cases of dyslexia are likely due to disorders in the visual modality. They are not further discussed here, as they are less well understood and they are of course not relevant for language genetics. Regarding theories of the phonological deficit as part of a pansensory disorder, we refer the reader to Ramus (2003).

REFERENCES

Abrahams, B. S., Tentler, D., Perederiy, J. V., Oldham, M. C., Coppola, G., & Geschwind, D. H. (2007). Genome-wide analyses of human perisylvian cerebral cortical patterning. *Proc. Natl. Acad. Sci. USA*, *104*(45), 17849–17854.

Adlard, A., & Hazan, V. (1998). Speech perception in children with specific reading difficulties (dyslexia). *Q. J. Exp. Psychol. [A]*, *51*(1), 153–177.

Alarcon, M., Abrahams, B. S., Stone, J. L., Duvall, J. A., Perederiy, J. V., Bomar, J. M., et al. (2008). Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am. J. Hum. Genet.*, *82*(1), 150–159.

Alcock, K. J., Passingham, R. E., Watkins, K. E., & Vargha-Khadem, F. (2000). Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain Lang.*, *75*(1), 17–33.

American Speech-Language-Hearing Association. (2007). Childhood apraxia of speech: Position statement [electronic version] from www.asha.org/policy.

Andrews, W., Barber, M., Hernadez-Miranda, L. R., Xian, J., Rakic, S., Sundaresan, V., et al. (2008). The role of Slit-Robo signaling in the generation, migration and morphological differentiation of cortical interneurons. *Dev. Biol.*, *313*(2), 648–658.

Anthoni, H., Zucchelli, M., Matsson, H., Muller-Myhsok, B., Fransson, I., Schumacher, J., et al. (2007). A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Hum. Mol. Genet.*, *16*(6), 667–677.

Arking, D. E., Cutler, D. J., Brune, C. W., Teslovich, T. M., West, K., Ikeda, M., et al. (2008). A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am. J. Hum. Genet.*, *82*(1), 160–164.

Bai, J. L., Ramos, R. L., Ackman, J. B., Thomas, A. M., Lee, R. V., & LoTurco, J. J. (2003). RNAi reveals doublecortin is required for radial migration in rat neocortex. *Nat. Neurosci.*, *6*(12), 1277–1283.

Bailey, J. A., Yavor, A. M., Viggiano, L., Misceo, D., Horvath, J. E., Archidiacono, N., et al. (2002). Human-specific duplication and mosaic transcripts: The recent paralogous structure of chromosome 22. *Am. J. Hum. Genet.*, *70*(1), 83–100.

Bakkaloglu, B., O'Roak, B. J., Louvi, A., Gupta, A. R., Abelson, J. F., Morgan, T. M., et al. (2008). Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am. J. Hum. Genet.*, *82*(1), 165–173.

Bartlett, C. W., Flax, J. F., Li, W., Reaple-Bonilla, T., Hayter, J., Hirsch, L. S., et al. (2003). A genome scan of specific language impairment loci in families from the United States. *Am. J. Hum. Genet.*, *73*(5), 491.

Bartlett, C. W., Flax, J. F., Logue, M. W., Vieland, V. J., Bassett, A. S., Tallal, P., et al. (2002). A major susceptibility locus for specific language impairment is located on 13q21. *Am. J. Hum. Genet.*, *71*(1), 45–55.

Bellini, G., Bravaccio, C., Calamoneri, F., Donatella Cocuzza, M., Fiorillo, P., Gagliano, A., et al. (2005). No evidence for association between dyslexia and DYX1C1 functional variants in a group of children and adolescents from Southern Italy. *J. Mol. Neurosci.*, *27*(3), 311–314.

Belton, E., Salmond, C. H., Watkins, K. E., Vargha-Khadem, F., & Gadian, D. G. (2003). Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. *Hum. Brain Mapp.*, *18*(3), 194–200.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behav. Brain Sci.*, *7*, 173–221.

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *J. Child Psychol. Psychiatry*, *31*(7), 1027–1050.

Bishop, D. V. M., Adams, C. V., Nation, K., & Rosen, S. (2005). Perception of transient nonspeech stimuli is normal in specific language impairment: Evidence from glide discrimination. *Appl. Psycholinguist.*, *26*, 175–194.

Bishop, D. V. M., Adams, C. V., & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes Brain Behav.*, *5*(2), 158–169.

Bishop, D. V. M., Bishop, S. J., Bright, P., James, C., Delaney, T., & Tallal, P. (1999). Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. *J. Speech Lang. Hear. Res.*, *42*(1), 155–168.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychol. Bull.*, *130*(6), 858–886.

Bonkowsky, J. L., & Chien, C. B. (2005). Molecular cloning and developmental expression of *foxP2* in zebrafish. *Dev. Dyn.*, *234*(3), 740–746.

Booth, J. R., Wood, L., Lu, D., Houk, J. C., & Bitan, T. (2007). The role of the basal ganglia and cerebellum in language processing. *Brain Res.*, *1133*(1), 136–144.

Brkanac, Z., Chapman, N. H., Matsushita, M. M., Chun, L., Nielsen, K., Cochrane, E., et al. (2007). Evaluation of candidate genes for DYX1 and DYX2 in families with dyslexia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, *144B*(4), 556–560.

Burbridge, T. J., Wang, Y., Volz, A. J., Peschansky, V. J., Lisann, L., Galaburda, A. M., et al. (2008). Postnatal analysis of the effect of embryonic knockdown and overexpression of candidate dyslexia susceptibility gene homolog Dcdc2 in the rat. *Neuroscience*, *152*(3), 723–733.

Carlsson, P., & Mahlapuu, M. (2002). Forkhead transcription factors: Key players in development and metabolism. *Dev. Biol.*, *250*(1), 1–23.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., et al. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*(5582), 851–854.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, *301*(5631), 386–389.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal behavior*. *Language*, *35*, 26–58.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clark, M. M., & Plante, E. (1998). Morphology of the inferior frontal gyrus in developmentally language-disordered adults. *Brain Lang.*, *61*(2), 288–303.

Colledge, E., Bishop, D. V., Koeppen-Schomerus, G., Price, T. S., Happe, F. G., Eley, T. C., et al. (2002). The structure of language abilities at 4 years: A twin study. *Dev. Psychol.*, *38*(5), 749–757.

Cope, N., Harold, D., Hill, G., Moskvina, V., Stevenson, J., Holmans, P., et al. (2005). Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am. J. Hum. Genet.*, *76*(4), 581–591.

De Fossé, L., Hodge, S. M., Makris, N., Kennedy, D. N., Caviness, V. S., Jr., McGrath, L., et al. (2004). Language-association cortex asymmetry in autism and specific language impairment. *Ann. Neurol.*, *56*(6), 757–766.

Dean, C., & Dresbach, T. (2006). Neuroligins and neurexins: Linking cell adhesion, synapse formation and cognitive function. *Trends Neurosci.*, *29*(1), 21–29.

Démonet, J.-F., Taylor, M. J., & Chaix, Y. (2004). Developmental dyslexia. *Lancet*, *363*(9419), 1451–1460.

Eckert, M. (2004). Neuroanatomical markers for dyslexia: A review of dyslexia structural imaging studies. *Neuroscientist*, *10*(4), 362–371.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*(6900), 869–872.

Felsenfeld, S., & Plomin, R. (1997). Epidemiological and offspring analyses of developmental speech disorders using data from the Colorado Adoption Project. *J. Speech Lang. Hear. Res.*, *40*(4), 778–791.

Feuk, L., Kalervo, A., Lipsanen-Nyman, M., Skaug, J., Nakabayashi, K., Finucane, B., et al. (2006). Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. *Am. J. Hum. Genet.*, *79*(5), 965–972.

Fisher, S. E. (2006). Tangled webs: Tracing the connections between genes and cognition. *Cognition*, *101*(2), 270–297.

Fisher, S. E., & DeFries, J. C. (2002). Developmental dyslexia: Genetic dissection of a complex cognitive trait. *Nat. Rev. Neurosci.*, *3*, 767–780.

Fisher, S. E., & Francks, C. (2006). Genes, cognition and dyslexia: Learning to read the genome. *Trends Cogni. Sci.*, *10*(6), 250–257.

Fisher, S. E., & Marcus, G. F. (2006). The eloquent ape: Genes, brains and the evolution of language. *Nat. Rev. Genet.*, *7*(1), 9–20.

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nat. Genet.*, *18*(2), 168–170.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, *97*(2), 179–210.

Flax, J. F., Realpe-Bonilla, T., Hirsch, L. S., Brzustowicz, L. M., Bartlett, C. W., & Tallal, P. (2003). Specific language impairment in families: Evidence for co-occurrence with reading impairments. *J. Speech Lang. Hear. Res.*, *46*(3), 530–543.

Fourcin, A. J. (1975a). Language development in the absence of expressive speech. In E. H. Lenneberg & E. Lenneberg (Eds.), *Foundations of language development* (Vol. 2, pp. 263–268). New York: Academic Press.

Fourcin, A. J. (1975b). Speech perception in the absence of speech productive ability. In N. O'Connor (Ed.), *Language, cognitive deficits and retardation* (pp. 33–43). London: Butterworths.

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage*, *20* (Suppl 1), S8–17.

Galaburda, A. M., & Kemper, T. L. (1979). Cytoarchitectonic abnormalities in developmental dyslexia: A case study. *Ann. Neurol.*, *6*(2), 94–100.

Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H., & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. *Nat. Neurosci.*, *9*(10), 1213–1217.

Galaburda, A. M., Sherman, G. F., Rosen, G. D., Aboitiz, F., & Geschwind, N. (1985). Developmental dyslexia: Four consecutive patients with cortical anomalies. *Ann. Neurol.*, *18*(2), 222–233.

Gauger, L. M., Lombardino, L. J., & Leonard, C. M. (1997). Brain morphology in children with specific language impairment. *J. Speech Lang. Hear. Res.*, *40*(6), 1272–1284.

Geschwind, N., & Galaburda, A. M. (1985). Cerebral lateralization. Biological mechanisms, associations, and pathology. I. A hypothesis and a program for research. *Arch. Neurol.*, *42*(5), 428–459.

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, *391*, 279–281.

Gopnik, M. (1990). Feature-blind grammar and dysphasia. *Nature*, *344*(6268), 715.

Gopnik, M., & Crago, M. B. (1991). Familial aggregation of a developmental language disorder. *Cognition*, *39*(1), 1–50.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., et al. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proc. Natl. Acad. Sci. USA*, *99*(16), 10911–10916.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the panglossian paradigm: A critique of the adaptationist programme. *Proc. R. Soc. Lond. B Biol. Sci.*, *205*(1161), 581–598.

Gray, P. A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., et al. (2004). Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science*, *306*(5705), 2255–2257.

Grigorenko, E. L. (2003). The first candidate gene for dyslexia: Turning the page of a new chapter of research. *Proc. Natl. Acad. Sci. USA*, *100*(20), 11190–11192.

Groszer, M., Keays, D. A., Deacon, R. M. J., de Bono, J. P., Prasad-Mulcare, S., Gaub, S., et al. (2008). Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr. Biol.*, *18*, 354–362.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of *FoxP2* in songbird basal ganglia nucleus Area X. *PLoS Biol.*, *5*(12), e321.

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., et al. (2004). *FoxP2* expression in avian vocal learners and non-learners. *J. Neurosci.*, *24*(13), 3164–3175.

Hallgren, B. (1950). Specific dyslexia (congenital word-blindness): A clinical and genetic study. *Acta Psychiatr. Neurol. Suppl.*, *65*, 1–287.

Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kääriäinen, H., et al. (2005). The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet.*, *1*(4), e50.

Harold, D., Paracchini, S., Scerri, T., Dennis, M., Cope, N., Hill, G., et al. (2006). Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. *Mol. Psychiatry*, *11*(12), 1085–1091, 1061.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.

Herbert, M. R., Ziegler, D. A., Deutsch, C. K., O'Brien, L. M., Kennedy, D. N., Filipek, P. A., et al. (2005). Brain asymmetries in autism and developmental language disorder: A nested whole-brain analysis. *Brain*, *128*(1), 213–226.

Herbert, M. R., Ziegler, D. A., Makris, N., Bakardjiev, A., Hodgson, J., Adrien, K. T., et al. (2003). Larger brain and white matter volumes in children with developmental language disorder. *Dev. Sci.*, *6*(4), F11–F22.

Humphreys, P., Kaufmann, W. E., & Galaburda, A. M. (1990). Developmental dyslexia in women: Neuropathological findings in three patients. *Ann. Neurol.*, *28*(6), 727–738.

Hurst, J. A., Baraitser, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Dev. Med. Child Neurol.*, *32*(4), 352–355.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends Cogn. Sci.*, *3*(7), 272–279.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*(2), 211–225.

Joseph, J. (2002). Twin studies in psychiatry and psychology: Science or pseudoscience? *Psychiatr. Q.*, *73*(1), 71–82.

Justus, T. (2004). The cerebellum and English grammatical morphology: Evidence from production, comprehension, and grammaticality judgments. *J. Cogn. Neurosci.*, *16*(7), 1115–1130.

Kaufmann, W. E., & Galaburda, A. M. (1989). Cerebrocortical microdysgenesis in neurologically normal subjects: A histopathologic study. *Neurology*, *39*(2, Pt. 1), 238–244.

Kovas, Y., & Plomin, R. (2006). Generalist genes: Implications for the cognitive sciences. *Trends Cogn. Sci.*, *10*(5), 198–203.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., et al. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr. Biol.*, *17*(21), 1908–1912.

Lai, C. S., Fisher, S. E., Hurst, J. A., Levy, E. R., Hodgson, S., Fox, M., et al. (2000). The SPCH1 region on human 7q31: Genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *Am. J. Hum. Genet.*, *67*(2), 357–368.

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*(6855), 519–523.

Lai, C. S., Gerrelli, D., Monaco, A. P., Fisher, S. E., & Copp, A. J. (2003). FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain*, *126*(Pt. 11), 2455–2462.

Lenneberg, E. H. (1962). Understanding language without ability to speak: A case report. *J. Abnorm. Soc. Psychol.*, *65*(6), 419–425.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Leonard, L. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Lewis, B. A. (1992). Pedigree analysis of children with phonology disorders. *J. Learn. Disabil.*, *25*(9), 586–597.

Liegeois, F., Baldeweg, T., Connelly, A., Gadian, D. G., Mishkin, M., & Vargha-Khadem, F. (2003). Language fMRI abnormalities associated with FOXP2 gene mutation. *Nat. Neurosci.*, *6*(11), 1230–1237.

Lopez-Bendito, G., Flames, N., Ma, L., Fouquet, C., Di Meglio, T., Chedotal, A., et al. (2007). Robo1 and Robo2 cooperate to control the guidance of major axonal tracts in the mammalian forebrain. *J. Neurosci.*, *27*(13), 3395–3407.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Ann. Dyslexia*, *53*, 1–14.

MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A. M., Lai, C. S., Vernes, S. C., et al. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am. J. Hum. Genet.*, *76*(6), 1074–1080.

Marcus, G. F. (2006). Cognitive architecture and descent with modification. *Cognition*, *101*(2), 443–465.

Marien, P., Engelborghs, S., Fabbro, F., & De Deyn, P. P. (2001). The lateralized linguistic cerebellum: A review and a new hypothesis. *Brain Lang.*, *79*(3), 580–600.

Marino, C., Giorda, R., Luisa Lorusso, M., Vanzin, L., Salandi, N., Nobile, M., et al. (2005). A family-based association study does not support DYX1C1 on 15q21.3 as a candidate gene in developmental dyslexia. *Eur. J. Hum. Genet.*, *13*(4), 491–499.

Marlow, A. J., Fisher, S. E., Francks, C., MacPhie, I. L., Cherny, S. S., Richardson, A. J., et al. (2003). Use of multivariate linkage analysis for dissection of a complex cognitive trait. *Am. J. Hum. Genet.*, *72*(3), 561–570.

Marshall, C. R., Harcourt-Brown, S., Ramus, F., & van der Lely, H. K. J. (submitted). Investigating phonological grammar in children with SLI and/or dyslexia: Is there compensation for place assimilation?

Marshall, C. R., Harcourt-Brown, S., Ramus, F., & van der Lely, H. K. J. (in press). The link between prosody and language skills in children with SLI and/or dyslexia. *Int. J. Lang. Commun. Disord.*

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the "specifics" of specific reading disability and specific language impairment. *J. Child Psychol. Psychiatry*, *41*(7), 869–874.

McBride, M. C., & Kemper, T. L. (1982). Pathogenesis of four-layered microgyric cortex in man. *Acta Neuropathol.*, *57*(2–3), 93–98.

Meng, H., Smith, S. D., Hager, K., Held, M., Liu, J., Olson, R. K., et al. (2005). DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc. Natl. Acad. Sci. USA*, *102*, 17053–17058.

Morley, M. E. (1967). *The development and disorders of speech in childhood*. Baltimore: Williams & Wilkins.

Nahon, J.-L. (2003). Birth of "human-specific" genes during primate evolution. *Genetica*, *118*(2–3), 193–208.

Newbury, D. F., Bonora, E., Lamb, J. A., Fisher, S. E., Lai, C. S., Baird, G., et al. (2002). FOXP2 is not a major susceptibility gene for autism or specific language impairment. *Am. J. Hum. Genet.*, *70*(5), 1318–1327.

Oliver, B. R., & Plomin, R. (2007). Twins' Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Res. Hum. Genet.*, *10*(1), 96–105.

Paracchini, S., Scerri, T., & Monaco, A. P. (2007). The genetic lexicon of dyslexia. *Annu. Rev. Genomics Hum. Genet.*, *8*, 57–79.

Paracchini, S., Thomas, A., Castro, S., Lai, C., Paramasivam, M., Wang, Y., et al. (2006). The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum. Mol. Genet.*, *15*(10), 1659–1666.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behav. Brain Sci.*, *13*(4), 707–784.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*(2), 201–236.

Plante, E. (1991). MRI findings in the parents and siblings of specifically language-impaired boys. *Brain Lang.*, *41*(1), 67–80.

Plante, E., Swisher, L., Vance, R., & Rapcsak, S. (1991). MRI findings in boys with specific language impairment. *Brain Lang.*, *41*(1), 52–66.

Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Curr. Opin. Neurobiol.*, *13*(2), 212–218.

Ramus, F. (2004). Neurobiology of dyslexia: A reinterpretation of the data. *Trends Neurosci.*, *27*(12), 720–726.

Ramus, F. (2006a). Genes, brain, and cognition: A roadmap for the cognitive scientist. *Cognition*, *101*(2), 247–269.

Ramus, F. (2006b). A neurological model of dyslexia and other domain-specific developmental disorders with an associated sensorimotor syndrome. In G. D. Rosen (Ed.), *The dyslexic brain: New pathways in neuroscience discovery* (pp. 75–101). Mahwah, NJ: Lawrence Erlbaum.

Ramus, F., Peperkamp, S., Christophe, A., Jacquemot, C., Kouider, S., & Dupoux, E. (in press). A psycholinguistic perspective on the acquisition of phonology. In C. Fougeron, B. Kühnert, & E. Delais-Roussarie (Eds.), *Papers in laboratory phonology X*.

Ramus, F., Pidgeon, E., & Frith, U. (2003). The relationship between motor control and phonology in dyslexic children. *J. Child Psychol. Psychiatry*, *44*(5), 712–722.

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Q. J. Exp. Psychol.*, *61*(1), 129–141.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., et al. (2006). Global variation in copy number in the human genome. *Nature*, *444*(7118), 444–454.

Rosen, G. D., Bai, J., Wang, Y., Fiondella, C. G., Threlkeld, S. W., LoTurco, J. J., et al. (2007). Disruption of neuronal migration by RNAi of Dyx1c1 results in neocortical and hippocampal malformations. *Cereb. Cortex*, *17*(11), 2562–2572.

Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *J. Phonetics*, *31*, 509–527.

Scerri, T. S., Fisher, S. E., Francks, C., MacPhie, I. L., Paracchini, S., Richardson, A. J., et al. (2004). Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK. *J. Med. Genet.*, *41*(11), 853–857.

Schumacher, J., Anthoni, H., Dahdouh, F., König, I. R., Hillmer, A. M., Kluck, N., et al. (2005). Strong genetic evidence for DCDC2 as a susceptibility gene for dyslexia. *Am. J. Hum. Genet.*, *78*, 52–62.

Serniclaes, W., Van Heghe, S., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *J. Exp. Child Psychol.*, *87*, 336–361.

Shriberg, L. D., Ballard, K. J., Tomblin, J. B., Duffy, J. R., Odell, K. H., & Williams, C. A. (2006). Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. *J. Speech Lang. Hear. Res.*, *49*(3), 500–525.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *J. Speech Lang. Hear. Res.*, *42*(6), 1461–1481.

Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., et al. (2005). Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proc. Natl. Acad. Sci. USA*, *102*(27), 9643–9648.

SLI Consortium. (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *Am. J. Hum. Genet.*, *70*(2), 384–398.

SLI Consortium. (2004). Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. *Am. J. Hum. Genet.*, *74*(6), 1225–1238.

Snowling, M. J. (2000). *Dyslexia* (2nd ed.). Oxford, UK: Blackwell.

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., et al. (2007). Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *Am. J. Hum. Genet.*, *81*(6), 1144–1157.

Stein, C. M., Millard, C., Kluge, A., Miscimarra, L. E., Cartier, K. C., Freebairn, L. A., et al. (2006). Speech sound disorder influenced by a locus in 15q14 region. *Behav. Genet.*, *36*(6), 858–868.

Stein, C. M., Schick, J. H., Gerry Taylor, H., Shriberg, L. D., Millard, C., Kundtz-Kluge, A., et al. (2004). Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *Am. J. Hum. Genet.*, *74*(2), 283–297.

Stephenson, S. (1907). Six cases of congenital word-blindness affecting three generations of one family. *Ophthalmoscope*, *5*, 482–484.

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, *315*(5813), 848–853.

Stromswold, K. (2001). The heritability of language: A review and meta-analysis of twin, adoption, and linkage studies. *Language*, *77*(4), 647–723.

Sun, T., Patoine, C., Abu-Khalil, A., Visvader, J., Sum, E., Cherry, T. J., et al. (2005). Early asymmetry of gene transcription in embryonic human left and right cerebral cortex. *Science*, *308*(5729), 1794–1798.

Szenkovits, G., Darma, Q., Darcy, I., & Ramus, F. (submitted). Exploring dyslexics' phonological deficit. II. Phonological grammar.

Taipale, M., Kaminen, N., Nopola-Hemmi, J., Haltia, T., Myllyluoma, B., Lyytinen, H., et al. (2003). A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc. Natl. Acad. Sci. USA*, *100*(20), 11553–11558.

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain Lang.*, *9*(2), 182–198.

Tallal, P. (2004). Improving language and literacy is a matter of time. *Nat. Rev. Neurosci.*, *5*(9), 721–728.

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.*, *29*(7), 382–390.

Tallal, P., Hirsch, L. S., Realpe-Bonilla, T., Miller, S., Brzustowicz, L. M., Bartlett, C., et al. (2001). Familial aggregation in specific language impairment. *J. Speech Lang. Hear. Res.*, *44*(5), 1172–1182.

Tallal, P., & Piercy, M. (1973). Developmental aphasia: Impaired rate of non-verbal processing as a function of sensory modality. *Neuropsychologia*, *11*(4), 389–398.

Tapia-Paez, I., Tammimies, K., Massinen, S., Roy, A. L., & Kere, J. (2008). The complex of TFII-I, PARP1, and SFPQ proteins regulates the DYX1C1 gene implicated in neuronal migration and dyslexia. *FASEB J*, *22*(8), 3001–3009.

Teichmann, M., Dupoux, E., Kouider, S., & Bachoud-Levi, A. C. (2006). The role of the striatum in processing language rules: Evidence from word perception in Huntington's disease. *J. Cogn. Neurosci.*, *18*(9), 1555–1569.

Temple, E. (2002). Brain mechanisms in normal and dyslexic readers. *Curr. Opin. Neurobiol.*, *12*(2), 178–183.

Teramitsu, I., & White, S. A. (2006). *FoxP2* regulation during undirected singing in adult songbirds. *J. Neurosci.*, *26*(28), 7390–7394.

Tomblin, J. B., & Pandich, J. (1999). Lessons from children with specific language impairment. *Trends Cogn. Sci.*, *3*(8), 283–285.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nat. Rev. Neurosci.*, *2*(10), 717–726.

van der Lely, H. K. J. (2005). Domain-specific cognitive systems: Insight from grammatical-SLI. *Trends Cogn. Sci.*, *9*(2), 53–59.

van der Lely, H. K. J., Rosen, S., & Adlard, A. (2004). Grammatical language impairments and the specificity of cognitive domains: Relations between auditory and language abilities. *Cognition*, *94*(2), 167–183.

van der Lely, H. K. J., Rosen, S., & McClelland, A. (1998). Evidence for a grammar-specific deficit in children. *Curr. Biol.*, *8*(23), 1253–1258.

Vargha-Khadem, F., Gadian, D. G., Copp, A., & Mishkin, M. (2005). FOXP2 and the neuroanatomy of speech and language. *Nat. Rev. Neurosci.*, *6*(2), 131–138.

Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., et al. (1998). Neural basis of an inherited speech and language disorder. *Proc. Natl. Acad. Sci. USA*, *95*(21), 12695–12700.

Velayos-Baeza, A., Toma, C., da Roza, S., Paracchini, S., & Monaco, A. P. (2007). Alternative splicing in the dyslexia-associated gene KIAA0319. *Mamm. Genome*, *18*(9), 627–634.

Velayos-Baeza, A., Toma, C., Paracchini, S., & Monaco, A. P. (2008). The dyslexia-associated gene KIAA0319 encodes highly N- and O-glycosylated plasma membrane and secreted isoforms. *Hum. Mol. Genet.*, *17*(6), 859–871.

Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., et al. (2008). A functional genetic link between distinct developmental language disorders. *N. Engl. J. Med.*, *359*, 2337–2345.

Vernes, S. C., Nicod, J., Elahi, F. M., Coventry, J. A., Kenny, N., Coupe, A. M., et al. (2006). Functional genetic analysis of mutations implicated in a human speech and language disorder. *Hum. Mol. Genet.*, *15*(21), 3154–3167.

Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., Davies, K. E., et al. (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *Am. J. Hum. Genet.*, *81*(6), 1232–1250.

Wang, Y., Paramasivam, M., Thomas, A., Bai, J., Kaminen-Ahola, N., Kere, J., et al. (2006). DYX1C1 functions in neuronal migration in developing neocortex. *Neuroscience*, *143*(2), 515–522.

Wassink, T. H., Piven, J., Vieland, V. J., Pietila, J., Goedken, R. J., Folstein, S. E., et al. (2002). Evaluation of FOXP2 as an autism susceptibility gene. *Am. J. Med. Genet.*, *114*(5), 566–569.

Watkins, K. E., Dronkers, N. F., & Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain*, *125*(Pt. 3), 452–464.

Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J., et al. (2002). MRI analysis of an inherited speech and language disorder: Structural brain abnormalities. *Brain*, *125*(Pt. 3), 465–478.

White, S., Frith, U., Milne, E., Rosen, S., Swettenham, J., & Ramus, F. (2006). A double dissociation between sensorimotor impairments and reading disability: A comparison of autistic and dyslexic children. *Cogn. Neuropsychol.*, *23*(5), 748–761.

White, S., Milne, E., Rosen, S., Hansen, P. C., Swettenham, J., Frith, U., et al. (2006). The role of sensorimotor impairments in dyslexia: A multiple case study of dyslexic children. *Dev. Sci.*, *9*(3), 237–255.

White, S. A., Fisher, S. E., Geschwind, D. H., Scharff, C., & Holy, T. E. (2006). Singing mice, songbirds, and more: Models for FOXP2 function and dysfunction in human speech and language. *J. Neurosci.*, *26*(41), 10376–10379.

Wigg, K. G., Couto, J. M., Feng, Y., Barr, C. L., Anderson, B., Cate-Carter, T. D., et al. (2004). Support for EKN1 as the susceptibility locus for dyslexia on 15q21. *Mol. Psychiatry*, *9*(12), 1111–1121.

Zeesman, S., Nowaczyk, M. J., Teshima, I., Roberts, W., Cardy, J. O., Brian, J., et al. (2006). Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. *Am. J. Med. Genet. A*, *140*(5), 509–514.

Zhang, J., Webb, D. M., & Podlaha, O. (2002). Accelerated protein evolution and origins of human-specific features: FOXP2 as an example. *Genetics*, *162*(4), 1825–1835.

59 The Biology and Evolution of Language: "Deep Homology" and the Evolution of Innovation

W. TECUMSEH FITCH

ABSTRACT The last decade has seen rapid and impressive progress in understanding the biology and evolution of complex "innovative" traits (e.g., insect wings or vertebrate eyes), and the fruits of this understanding are beginning to have an impact on our understanding of that most innovative of human traits: language. Although language, as a whole, is unique to *Homo sapiens*, many of the neural and cognitive mechanisms supporting language are shared with other species. An empirically based, mechanistic understanding of the evolution of language therefore requires research on both unique aspects of language (such as complex syntax) and broadly shared features. Evolutionary developmental biology ("evo-devo") has added a new twist to this distinction, with the discovery that traits shared due to *convergent* evolution (such as vocal learning in humans and birds) may nonetheless be based on *homologous* genes and developmental pathways. Such "deep homologies" may involve convergence at the phenotypic level and homology at the genotypic level, and illustrate the need to rethink traditional ideas about homology. Studies of eyes, limbs, and body plans have revealed deep homologies in all these systems. Here, I suggest that language is also likely to have its share of deep homologies, and that this possibility provides a powerful rationale for investigations of convergently evolved traits in widely separated species. I illustrate the potential of this new approach with an exploration of the neural and genetic basis of vocal learning in humans and birds. I conclude that neuroethological investigations of diverse vertebrate species, from fish to birds to mice, will powerfully augment more traditional work on primates in the search for the neural mechanisms underlying language.

Humans (like most species) are unique in many ways, but language is the jewel in our cognitive crown. Language makes possible the greatest human cultural achievements, ranging from quantum physics to the novel to the Internet, because knowledge can be conveyed from mind to mind, across generations, with progressive refinements and elaboration. Without language and the community of minds that it creates, our species would be little more than an unusually clever bipedal ape. It is clear that human language rests on a unique, recently evolved, biological basis: our nearest living relatives, the chimpanzees, are unable to acquire language past the level of a young child. Sometime in our recent evolution, in the last 5–7 million years since we diverged from our last common ancestor with chimpanzees, a suite of important innovations have occurred, which together comprise the human capacity to acquire language. The nature of, and biological basis for, this capacity is a core focus of contemporary research on the biology of language.

Languages themselves, like English or Chinese, are obviously not inborn. We each acquire the language of our local community through an experience-dependent process of language acquisition. In Darwin's words "language is an art . . . not a true instinct, for every language has to be learnt. It differs, however, widely from all ordinary arts, for man has an instinctive tendency to speak, as we see in the babble of our young children; whilst no child has an instinctive tendency to brew, bake, or write" (Darwin, 1871, p. 55).

Today there is wide agreement that the language acquisition process has a strong biological basis and represents an "instinct to learn" that is part of every normal child's genetic heritage. Although this human capacity is unique when considered as a whole (sometimes termed the "faculty of language in a broad sense," or FLB), most of the component mechanisms underlying language are not unique to our species. Factors shared with chimpanzees and other primates (e.g., mechanisms underlying lexical acquisition) are traditionally believed to be homologies, traits that were present in our shared common ancestor. Other traits are shared with more distant biological relatives but *not* with other primates (e.g., mechanisms underlying imitative vocal learning). Such traits are traditionally considered to represent convergent evolution, analogy, or "homoplasy."

This latter category is of particular interest to neuroscientists, because animal species sharing linguistically relevant traits like vocal learning (e.g., songbirds) are more amenable to experimental analysis than are chimpanzees or other nonhuman primates. It is becoming increasingly clear that traits that have evolved independently ("mere" analogies) may nonetheless be based on shared genetic developmental

W. TECUMSEH FITCH School of Life Sciences, University of Vienna, Vienna, Austria

pathways: a situation sometimes termed *deep homology*. Recent discoveries suggest that deep homology is common in evolution (Shubin, Tabin, & Carroll, 1997; Wilkins, 2002). In deep homologues the genetic pathways leading to a trait are shared, even though the structures that they build are not necessarily themselves homologous, because they were absent in phylogenetically intermediate taxa. Examples include limbs in vertebrates and crustaceans, wings in birds and insects, and the eye in vertebrates and cephalopods (squids or octopuses).

The discovery of deep homology provides an exciting new range of empirical possibilities for scientists interested in the evolution of complex innovations, including human language. The very concept of deep homology would have been considered fanciful 20 years ago, and its reality has profound consequences for both the concept of homology and our understanding of the evolution of complex innovations. The purpose of this article is to explore some of the consequences of this new understanding for human language.

Understanding human cognitive evolution

THE GENETIC CHALLENGE OF LANGUAGE Humans are unusual animals in many respects. Our most prominent morphological differences from our nearest extant relatives, the chimpanzees, are upright posture and bipedal locomotion, relatively large brains, small teeth, and relative hairlessness, as well as a host of cognitive differences including sophisticated tool use and language. What are the genetic determinants of these differences? Now that both human and chimpanzee genomes have been sequenced, and given that there is only roughly 1.2% genetic divergence at the sequence level (Chen & Li, 2001), one might expect this question to be an easy one to answer. Unfortunately, this is far from being the case: the human genome contains roughly 3×10^9 base pairs, predicting a whopping 18 *million* changes during our recent evolution (assuming that chimpanzees and humans each account for half of the difference from a common ancestor, and thus that 0.6% of the changes occurred in the human genome). Many of these changes will be "silent" in the sense of having no effect on the phenotype, constituting "noise" in our search for the underlying cause of human-specific traits (Carroll, 2003). Thus only a small proportion of these 18 million changes will be biologically meaningful, and an even smaller proportion may be expected to correlate specifically with human cognitive differences. Finding out which genes these are and what specific changes led to which phenotypic effects is a great challenge facing our attempts to understand the genetic mechanisms underlying human evolution.

The challenge is made greater by the fact that, based on what is known from "model organisms" such as mice or *Drosophila*, many of the critical changes are expected to involve the expression of regulatory genes and noncoding DNA regions to which these genes bind, rather than protein-coding genes (Carroll, 2003). Protein-coding genes account for a modest 1.5% of the human genome, and many base-pair changes are synonymous, so only some 200,000 base-pair changes (about 1% of the total) will lead to protein differences. Current techniques for finding such coding differences are well developed and allow us to discover genes that have been subjected to selection (e.g., Fay, Wyckoff, & Wu, 2001). Unfortunately, the noncoding portion of the genome remains far more difficult to analyze. Differences in gene expression between humans and other nonhuman primates reveal substantial differences in the levels at which the same genes are expressed in the brain and suggest that humans are unusually divergent in this respect (Enard, Khaitovich, Klose, & Paäbo, 2002). Thus we have good reasons to believe that changes in gene regulation and in noncoding regions of DNA will be central to understanding the evolution of human cognitive capabilities including language. Given that many of the genetic differences between us and chimpanzees have to do with immunity, olfaction, and reproduction (Chimpanzee Sequencing and Analysis Consortium, 2005), finding and understanding the genes underlying human cognitive capacities resembles the proverbial search for a needle in a haystack. This is the central challenge that needs to be faced if progress is to be made understanding the genetic basis of language. Researchers hoping to meet this challenge will need all the help they can get, from all the disciplines potentially involved.

THE NEURAL CHALLENGE OF LANGUAGE The evolution of language has been argued to be one of the "hardest problems in science" (Christiansen & Kirby, 2003), because its solution requires progress on so many different disciplinary fronts (including neuroscience, linguistics, psychology, and evolutionary biology). Real solutions will require successful interdisciplinary interactions across these traditionally separate scientific domains. For the cognitive neuroscientist, the fact that language enables us to express virtually any aspect of our thoughts and feelings suggests that any aspect of conscious brain function, across many cortical areas, may potentially be tapped for expression during linguistic encoding. Exclusive focus on Broca's area or other traditional "language" areas will not by itself be sufficient. Language takes all of cognition as its potential domain, and a well-developed neural theory of cognition will be required for a full understanding of the neural basis of language. Such a theory is still far off. Furthermore, it is now clear that much of the neural machinery for language is epigenetically specified (there is no "hard-coded" language processor in a fixed brain region). Adequate language skills develop even after complete removal of the left hemisphere (Liégeois et al., 2004), and bilinguals can recruit different brain regions when processing different languages (Hull & Vaid, 2007).

Thus an adequate neural model will also need to incorporate a rich theory of neural epigenesis: the interactions between gene expression in independent cells as influenced by their local neural processing environment within the brain and the external sensory world. This understanding also appears far off, though recent progress driven by molecular techniques is cause for hope. It will require the abandonment of *scala natura* models of brain evolution (cf. Striedter, 2004), and accepting the broad and deep similarities among *all* vertebrate brains, from fish to humans, while simultaneously allowing for the equally important differences among the brains of even closely related species.

The Linguistic Challenge For the (psycho)linguist, the cognitive revolution in psychology and the associated generative revolution in linguistics have led to both considerable progress and a confusing profusion of theoretical frameworks and perspectives. While there is widespread agreement that the capacity for language has a strong biological basis, unique when considered as a whole to our species, there is little consensus about the detailed nature of this capacity. Plausible hypotheses include a continuum from a broad and general "capacity for culture" (Tomasello, 1999; Tomasello, Carpenter, Call, Behne, & Moll, 2005) to a detailed computational system specific to language and to humans (Pinker & Jackendoff, 2005). Intermediate positions include the possibility that language is built on a broadly shared cognitive foundation, with a few powerful but novel computational operations knitting shared mechanisms together (Hauser, Chomsky, & Fitch, 2002; Fitch, Hauser, & Chomsky, 2005). We can refer to *all* of the cognitive/neural mechanisms involved in language production, perception, or processing as the "faculty of language in a broad sense," or FLB; the subset of processing mechanisms specific to language and to our species can then be referred to as the "faculty of language in a narrow sense," or FLN (Hauser et al.). Because the nature of this latter, more specific, subset remains highly controversial (Fitch et al.; Pinker & Jackendoff, 2005), I conservatively take the FLB as a whole as the explanandum here, emphasizing that this is a multicomponent system. The FLB includes the mechanisms underlying phonetics, phonology, syntax, semantics, and pragmatics, without regard to whether each component is specific to humans or not. Many components of the FLB will be shared with nonhuman animals and thus can be studied comparatively, and such studies are required to determine the contents of FLN.

The Evolutionary Challenge For the evolutionary biologist, language evolution raises a number of challenges as well. One is that the cultural transmission of language and the resultant process of language change ("glossogeny") raise important issues beyond those raised by the interaction of ontogeny and phylogeny for all traits. Because language changes, the linguistic target of the learner has been filtered through the minds of previous humans, and this process leads to an "evolutionary" dynamic of its own (Darwin, 1871; Fitch, 2007b; Kirby, Smith, & Brighton, 2004; Pagel, Atkinson, & Meade, 2007). The importance of this interposition of an additional form of glossogenetic change, with a time scale between that of ontogeny and phylogeny, is increasingly recognized (Deacon, 1997; Hurford, 1990; Kirby et al.; Nettle, 1999), but scientists are only beginning to resolve some of the problems in gene/culture coevolution that this raises (Richerson & Boyd, 2005). Fortunately, we have other biological examples of cultural change for comparison (e.g., bird or whale "song").

The same is not true for the central feature of semantics: the ability to express arbitrary thoughts. Current understanding of animal communication strongly suggests that humans are unique in this ability: if there are nonhuman species with open-ended semantics, they are remarkably clever at hiding these abilities from generations of dedicated ethologists (Bradbury & Vehrencamp, 1998; Hauser, 1996). Efforts at training nonhuman species with languagelike systems reveal both commonalities and significant differences (Kako, 1999; Tomasello & Call, 1997). One of the sharp differences is the apparent drive in our species to express our thoughts and feelings to others. This drive poses significant evolutionary problems, for the evolution of such apparently "altruistic" behavior is not predicted by standard models of the evolution of communication and cooperation (Axelrod & Hamilton, 1981; Dawkins & Krebs, 1978; Trivers, 1971; Zahavi, 1993), while the kin-selection route to cooperative communication (Hamilton, 1964; Maynard Smith & Harper, 2003) appears confounded by the fact that humans do not communicate exclusively with kin (cf. Fitch, 2004). Thus language evolution still poses deep evolutionary puzzles, if not the "embarrassment to evolutionary theory" once proclaimed by Premack (1986).

This brief survey should convince any skeptics that the biology and evolution of language involve a profusion of challenging scientific problems. For many years the topic of language evolution was neglected as a result. However, a series of methodological advances have combined with theoretical progress in a number of disciplines to reawaken hope that "biolinguistics," as this field is sometimes called, can become an productive, empirical scientific discipline. This hope is exciting, because language is such a central aspect of human nature that a failure to understand it entails a fundamentally incomplete understanding of ourselves. I optimistically believe that the challenges I have sketched, though daunting, are within the realm of scientific inquiry and will eventually yield to concerted empirical research. Indeed, I think that with collaborative interdisciplinary effort

many of them can be solved in this century. One reason for optimism is the success of the new synthesis of evolutionary and developmental biology, which has revolutionized contemporary understanding of the interactions of genes, development, and evolution in the last decades.

"Evo-devo": The recent marriage of evolutionary and developmental biology

At an interdisciplinary conference on adaptation and evolution in 1979, the great evolutionary biologist John Maynard Smith remarked that evolutionary theory will remain incomplete until integrated with a similarly advanced theory of developmental biology. Because it is genes, not individuals, that are passed down across the generations, mathematical approaches to evolutionary biology take changes in *genotypes* as the core phenomenon of interest (specifically, population-level changes in allele frequency), and yet Darwin's theory focuses on variant *phenotypes* (the material form of individual organisms). Clearly, a bridging theory between genotype and phenotype is required: molecular developmental biology and a successful theory of ontogenesis. Today, by means of a happy series of consilient advances, the foundations of such a theory are now available in a discipline called *evolutionary developmental biology*, or "evo-devo" (Carroll, 2005; Carroll, Grenier, & Weatherbee, 2005; Wilkins, 2002). Central to this discipline is the recognition that genes do not direct development in a vacuum, but that gene expression within a cell always occurs in an environment made up of other cells and structured by the past behavior of previous cells. Genes do not provide a blueprint of adult form, but rather a development "recipe" that relies heavily on the local extracellular environment to determine the differentiation and growth of individual, semi-independent cells. This continual process of interaction between DNA within the cell and the extracellular environment is termed *epigenesis*, a term connoting the inadequacy of genetic determinism and traditional "gene as blueprint" models of ontogeny.

Neural Epigenesis Although our understanding of the epigenetic cycles of interaction between cells and their local environment is currently most advanced for morphological features such as animal limbs (Carroll et al., 2005), an epigenetic perspective is equally crucial for understanding brain development (Nottebohm, 1989; Rakic, 1985). This is particularly true of the vertebrate neocortex, characterized by extreme plasticity during fetal development. For example, temporal cortex, normally slated to function as "auditory cortex" in the adult, can be "rewired" to serve as visual cortex early in ontogeny (von Melchner, Pallas, & Sur, 2000). Such data extend the long-appreciated dependence of the developing brain on organism-external stimulation (Held & Hein, 1963; Hubel & Wiesel, 1965) to the organism-internal "environment" that influences cellular differentiation. The importance of this plasticity, though greatest in early development, never ceases in the vertebrate brain (Nottebohm, 1989; Nottebohm, Kasparian, & Pandazis, 1981; Schlaug, Jäncke, Huang, & Steinmetz, 1995). Thus the epigenetic perspective of evo-devo is just as pertinent to understanding the brain as for the development of the hand or the eye (Carroll, 2003; Fernald, 2000; Shubin et al., 1997). Under this broad epigenetic umbrella, one of the most crucial recent advances has been the discovery of deep conservation of genetic mechanisms and pathways involved in development.

"Deep homology" and the conservation of developmental pathways

The great evolutionist Ernst Mayr once wrote that "the search for homologous genes is quite futile except in very close relatives" (Mayr, 1963, p. 609). A recent revolutionary finding of modern molecular biology is that Mayr was quite wrong about this. We now know that genes, particularly those involved in development, can be conserved over extremely long periods of evolutionary time (e.g., the 1200 million years separating contemporary arthropods and vertebrates). This discovery occurred initially in the context of HOX genes ("homeobox" genes, which play a crucial role in patterning the anterior-posterior body axis; cf. Duboule, 1994; Lewis, 1978). However, it is now apparent that many other developmental genes are equally strongly conserved (Wilkins, 2002), especially those that code for transcription factors (proteins that bind to DNA and influence subsequent gene expression) and their binding sites. Further examples include the Pax6 gene involved in the specification of eyes in diverse species from flies to squid to mice to humans (Gehring & Ikeo, 1999; Tomarev et al., 1997; van Heyningen & Williamson, 2002).

What makes the existence of such shared genes and genetic pathways (Duboule, 2007) deeply surprising is that they control the development of structures that evolved convergently. For example, the wings of birds and flies evolved independently: the common ancestor of insects and birds did not posess wings (it was a wormlike terrestrial organism sometimes known as the Ur-bilaterian; De Robertis & Sasai, 1996). Fly and bird wings are thus "analogues" rather than homologues. It is thus surprising that the same underlying genetic mechanisms build these structures. This example is not particularly troubling, because in both cases we can consider the wings to be evolved from simple generalized appendages that were present in the Ur-bilaterian. The case of Pax6 in eye development is more significant, because virtually all pre-evo-devo scholars interested in eye evolution agreed that complex eyes have evolved convergently in insects, molluscs, and vertebrates. Indeed the camera eye,

with lens and retina, of vertebrates and cephalopods (squid and octopus) is a textbook case of convergent evolution. But we now know that Pax6 controls eye development in both. We must thus hypothesize that some simple eyespots in the Ur-bilaterian were, in some sense, ancestral to modern complex eyes, but this hypothesis greatly stretches the traditional morphological definition of homology.

The solution to this problem has been to recognize a new possibility, that *developmental programs*, down to the detailed level of gene sequences and expression patterns during development, may be shared by virtue of common descent (and thus homologous in one sense) while the *structures that they build* are not. From a structural viewpoint, squid and human eyes are convergently evolved analogues, but from a developmental viewpoint the genetic tools *utilized* to build them are homologues. This superficially paradoxical situation results from deep homology, and it demands wholesale rethinking of traditional notions of homology (Rutishauser & Moline, 2005; Shubin et al., 1997). Given increasing evidence that developmental pathways are highly conserved, across all metazoan phyla, deep homology may be common and indeed may be the rule rather than the exception in development. This evidence raises the possibility that shared homologous developmental pathways underlying the evolution of such innovative traits as eyes, wings, limblessness in reptiles, or echolocation in bats might have implications for debates in the biology and evolution of language.

Deep homology and evolutionary innovation in cognition and language

A persistent debate in cognitive science, inherited by cognitive neuroscience, may be characterized as the "specialist/generalist" debate. At the "specialist" end of the continuum, often typified by neuroethologists, organisms are seen as supremely adapted to their particular way of life: echolocating bats have evolved specializations of hearing and vocal production, electroreceptive fish have evolved innovative electrical field production and perception mechanisms, and food-caching birds have evolved prodigious memories (Camhi, 1984; Schnitzler, Menne, Kober, & Heblich, 1983). However, this perspective has not gone unchallenged by researchers who point out that underlying mechanisms may be shared across superficially different cognitive domains (cf. Bolhuis & Macphail, 2001). Similarly, a dominant contemporary paradigm in human evolutionary psychology favors a view of the human mind/brain as a "Swiss army knife": a series of domain-specific adaptive modules (Barkow, Cosmides, & Tooby, 1992). Opponents of this nativist/modularist view emphasize the deep similarities in cognitive performance both across cognitive domains and between species (for a balanced review see Laland & Brown, 2002). We might characterize these different viewpoints as emphasizing "modularist specialization" versus "generalist universality." Both viewpoints have been applied to human language.

As cogently observed by Heyes (2003), these different viewpoints about cognitive evolution need not be in opposition. A multicomponent perspective on language suggests that some components (e.g., long-term memory for lexical items) might be broadly shared between species and cognitive domains, while others (e.g., mechanisms underlying recursive syntax) might be unique to our species (Hauser, Chomsky, et al., 2002). Such a mixed bag is expected from Darwinian evolution, and a theoretical framework for understanding language evolution must fully encompass both possibilities if we are to empirically resolve the issue. Arguments that "language" is monolithically modular or domain-general both oversimplify the situation to the detriment of empirical progress. Instead, humans have multiple mechanisms (biologically based predilections, biases, and constraints) crucial to language. Once we have accurately specified the particular mechanism of interest (e.g., vocal learning, syntax comprehension, lexical acquisition, "theory of mind," etc.), it becomes an empirical question whether the components underlying such traits constitute widely shared mechanisms making up a general vertebrate or mammalian "cognitive tool kit" or highly specific components uniquely tuned to human language. Probably, some will be shared, and some will be unique to our species. We expect even "unique" mechanisms to function in a context of a suite of shared cognitive mechanisms that both predated them in evolutionary time and are shared with nonhuman species. A broad comparative approach is a logical prerequisite for addressing such questions, because no valid claim of "human uniqueness" can be made without a search for similar mechanisms in other animals.

The discovery of deep homology raises the fascinating possibility that even "unique" innovations, evolved during recent human evolution and isolated to our small branch of the primate lineage, might derive from more widely shared developmental processes. If so, the developmental pathways involved may be expected to impose certain constraints (or biases) on the system thus evolved, constraints that can be understood by examining the nature of the developmental process in nonhuman species. This approach provides an exciting empirical possibility: that the nature of the neural developmental processes that give humans our unique capacity for language can be probed, at a detailed molecular genetic level, by examining analogous processes in other vertebrates. Indeed, such inquiry could actually aid in *gene discovery* and thus help solve the "needle in a haystack" problem discussed earlier. To the extent that nature repeatedly uses the same developmental tool kit to solve similar evolutionary problems, generating deep homology, we can expect investigations of widely separated organisms, from honeybees to birds, to offer valuable cues to the genes and

genetic pathways involved in human language. This endeavor requires that we specify the nature of the computations that a particular linguistic mechanism performs, and then search for analogues of these computations in other species, regardless of how closely related to us they might be. Although we don't expect this strategy to work in every case (no one expects that *every* convergently evolved trait will involve deep homology), the recent evo-devo literature suggests that deep homology is common enough to constitute a winning bet in many cases (Carroll, 2006; Carroll et al., 2005; Shubin, 2008; Wilkins, 2002). Thus the concept of deep homology opens exciting routes for empirical enquiry.

In the remainder of this chapter I review a specific set of neural and genetic mechanisms, relevant to speech, from a comparative evolutionary viewpoint. These data illustrate the reality of a deep homology underlying vocal learning, demonstrating the potential for this approach to the biology and evolution of human language more generally.

An illustrative example: Mechanisms underlying speech and vocal learning

Among primates, humans appear unique in our capacity for complex vocal learning, but we share this capacity with dolphins, seals, and birds (Fitch, 2000; Janik & Slater, 1997) and probably elephants (Poole, Tyack, Stoeger-Horwath, & Watwood, 2005). It has recently become clear that vocal learning in birds and humans nonetheless involves identical genes, and in particular that the FOXP2 gene involved in human oromotor control (see chapter 58 in this volume by Franck Ramus and Simon E. Fisher) also plays a crucial role in vocal learning in zebra finches (Haesler et al., 2007; Scharff & Haesler, 2005). Thus FOXP2, the first gene known to be specifically involved in human spoken language, also appears to be a case of deep homology, in the same vein as the well-studied examples of HOX genes in body plan development or PAX6 in eye development.

A core prerequisite of a flexible communication system is that it include a mechanism for innovation (of new words, phrases, or syntactic rules) and a way of sharing these among those using the language. These requirements entail a capacity for imitation: children learning this system must be able to copy lexical items or rules that are not part of their innate behavioral repertoire. Because language can be expressed either vocally-auditorily (speech) or visually-manually (sign), this prerequisite is a general one. However, it is clear that great apes, and thus our common ancestor with chimpanzees, have a poor or nonexistent capacity for vocal imitation, while their visual/gestural imitation is more like our own (Call & Tomasello, 2007; Hayes, 1951; Janik & Slater, 1997). Thus a "key innovation" (Liem, 1973) in the evolution of spoken language was our novel capacity for vocal imitation.

A key component of vocal motor control, present in humans but lacking in other primates, is the direct connections between motor neurons in the inferior frontal neocortex to the motor neurons in the brain stem that control the larynx and respiration (Deacon, 1997; Jürgens, 2002; Kuypers, 1958a, 1958b). Mammalian vocalization, including innate vocalizations like screams and laughter in humans, generally relies upon an ancient medial "limbic" system, incorporating command neurons in the periacqueductal gray (PAG) that project to the medullary nuclei that are home to vocal motor neurons, particularly the nucleus ambiguus. While this ancient "chassis" controlling vocalization includes a medial cortical component (the anterior cingulate), neurons in the lateral cortex play little role in controlling vocalization and lack any direct connections onto vocal motor neurons in most vertebrates (Jürgens, 2002, 1994). Humans have evolved an additional set of vocal connections, constituting a novel system for vocal control, parallel to this ancient mammalian system. Kuypers and Jürgens hypothesized that this system, with its direct connections from motor cortex to vocal motor neurons, is a crucial component in the human ability to voluntarily control the acoustic details of vocalization: an ability not present in monkeys or apes (Larson, Ortega, & DeRosier, 1988; Larson, Sutton, Taylor, & Lindeman, 1973). Consistent with this hypothesis, no such direct connections have been revealed in detailed studies of many different nonhuman primates (reviewed in Jürgens, 2002).

This *Kuypers/Jürgens hypothesis*—that direct corticomotor connections are necessary for complex vocal control—is testable by examinations of nonprimate species in which vocal learning is or is not present. For example, vocal learning has independently evolved in three clades of birds, all of which show surprisingly similar patterns of connectivity that differentiate them from nonvocal learners (Jarvis, 2007). The best-studied of these three groups, by far, are the oscine passerine birds ("songbirds") all of whom appear to learn their songs. Consistent with the Kuypers/Jürgens hypothesis, direct telencephalic/vocal-motor neuron connections exist in songbirds and parrots (Jarvis, 2004; Wild, 1993) comparable to those documented in our own species. This finding illustrates the power of the comparative method to use convergently evolved traits to test adaptive, mechanistic hypotheses.

The connections discussed in the previous paragraph are probably necessary for vocal control but not alone sufficient for vocal learning, which further requires a capacity to utilize auditory input to control vocal output. This process, too, is well understood in songbirds (Marler & Slabbekoorn, 2004), often relying on a developmental process in which the young bird is exposed to song during a sensitive period, and then fine-tunes its song and its control over the vocal apparatus during a practice period termed *subsong*. Songbirds deprived of either "template" songs or the opportunity for subsong

do not typically develop species-typical song. The subsong phenomenon in birds provides a striking parallel with human babbling, which appears to play a similar functional role in human vocal learning (Doupe & Kuhl, 1999; Locke & Pearson, 1990). This, then, is a shared behavioral mechanism, similar to the neural mechanisms just discussed, that underpins the convergence of vocal learning in humans and birds.

Finally, recent genetic studies of birdsong learning demonstrate further similarities. Recent experimental work in the laboratory of Constance Scharff has now clearly documented a role in avian vocal learning of a gene that was originally discovered in the context of human vocal motor control and learning: forkhead-box P2, or FOXP2 (cf. Ramus and Fisher, chapter 58 in this volume). Like the HOX and PAX genes discussed earlier, FOX genes code for a transcription factor: a protein that binds to DNA and enhances or inhibits the expression of other genes. Also like HOX and PAX, FOX genes are members of a large and highly conserved family of transcription factors. Perhaps surprisingly, given this conservatism, a human-specific mutation in this gene leads to a specific deficit in oral and vocal motor control, first discovered in a family living in England (Vargha-Khadem, Gadian, Copp, & Mishkin, 2005; Vargha-Khadem & Passingham, 1990; Vargha-Khadem et al., 1998). The discovery of this gene (Fisher, Vargha-Khadem, Watkins, Monaco, & Pembrey, 1998) was groundbreaking in that it uncovered the first, presumably of many, genes involved in human cognition and language. A decade later FOXP2 remains the clearest example of a gene involved in spoken language, shared by all nonclinical human populations, and different from chimpanzees and other primates.

In a striking new demonstration of deep homology, Scharff and her colleagues have shown that FOXP2 (and other closely related FOX genes, including FOXP1) is expressed in similar brain regions in birds and humans and plays a role in vocal learning (Haesler et al., 2004; Scharff & Haesler, 2005). In the most direct evidence of this role, Haesler and colleagues (2007) showed that a novel lentivirus-mediated knockdown of FOXP2 expression, via RNA interference, decreased the quality and quantity of vocal learning in zebra finches. The effect occurred only with injections in brain regions specifically evolved in vocal learning (Area X, homologous to basal ganglia in humans) and not injections in nonsong areas, providing strong evidence for a key role of FOXP2 in vocal learning in birds. Although the ability for vocal learning evolved separately in birds and humans, the behavioral and neural mechanisms involved show that there are fundamental similarities at the computational and circuit levels, and that the genetic mechanisms involved are identical. Thus FOXP2 constitutes a deep homology: a conserved homologous developmental pathway underlying a convergently evolved trait.

Although there are presumably many genes involved in vocal motor control, and many more involved in language more broadly, the example of FOXP2 shows that deep homology is not a phenomenon restricted to peripheral morphology. Furthermore, it illustrates the power of a model system like birdsong to illuminate our understanding of human vocal control. Experiments like those just discussed in birds are impossible in humans for ethical reasons and in primates for practical reasons: primates are incapable of complex vocal learning. While FOXP2 knockout mice have been created that show various motor deficits (Shu et al., 2005) and knockin mice with human versions of FOXP2 have been engineered (Groszer et al., 2008), there is no evidence of vocal learning in mice, and thus interpretation of these results will remain problematic. Indeed, the developmental processes in which FOXP2 is involved almost certainly require a suite of other coevolved mechanisms that are not present in most mammals. In contrast, in species like songbirds with a fully developed vocal learning ability, the discovery of one of the genes involved opens the door to targeted search for other genes in the system. The use of large-scale gene-expression assays, targeted on genes known to be up-regulated during vocal learning in birds (Wada et al., 2006), will play an important role in the discovery of such genes. Thus the undisputed fact that birdsong and human speech evolved independently may turn out to be quite irrelevant to the question of the genetic mechanisms involved, which may well be largely homologous.

Conclusions and prospects

In this chapter I have argued that the discovery of deep homology is relevant to cognitive neuroscience, and in the case of FOXP2 to spoken language. However, I fully appreciate that speech is not language (Fitch, 2000) and constitutes just one component of a set of diverse mechanisms necessary for human language. What of these other mechanisms? In particular, what of semantics and syntax, which most scholars agree are more central to human language than is speech (though see Lieberman, 1998, 2000)? At present we know far less about the neural and genetic mechanisms underlying semantics or syntax than those underlying speech, but a combination of brain imaging, gene expression profiling, and exploitation of the comparative method gives reasons for optimism concerning these components of language. I will thus end by listing some open questions concerning these additional factors.

Semantics A central challenge language poses for evolutionary theory is the readiness humans exhibit to share information with other, unrelated, individuals. This drive is striking in its absence in most animals, even in language-trained chimpanzees who have the machinery for transmitting

considerable information about the world (Tomasello, 2001, 2007). Although our loquacious tendency has no specific name in English, the German term *Mitteilungsbedürfnis* (denoting a drive to share one's thoughts) captures it nicely. This drive is puzzling to evolutionary biologists because the standard explanation of apparently altruistic behaviors in animals is that they are genetically "selfish" in the sense that they are doled out preferentially to related individuals, and can thus be explained by means of kin selection (Foster, Wenseleers, & Ratnieks, 2006; Maynard Smith, 1964). But humans share information not only with unrelated individuals, but also with perfect strangers, a practice for which there can be little evolutionary advantage. One way out of the conundrum is the suggestion that language evolved initially in a kin communication context and only later was "exapted" out of this context to serve unrelated individuals exchanging information reciprocally (Fitch, 2004, 2007a). By this model, the friendly sharing of information with total strangers might remain a nonadaptive characteristic of human language. Alternatively, this may have supported the evolution of novel mechanisms for more Machiavellian reciprocity among unrelated adults. These issues remain largely unstudied empirically.

Another deep problem in the evolution of semantics is the evolution of "Theory of Mind" (ToM) in our species. Many aspects of the pragmatics of normal conversation require not just a *desire* to be informative, but also a complex mental model of what one's interlocutor knows and does not know (Grice, 1975). Being informative in the simple sense of relaying new information is not enough: to be *relevant* also requires a further notion of what the other person is trying to accomplish in a particular conversation (Sperber & Wilson, 1986). My mother's birthday or grandfather's shoe size are likely to be novel to most interlocutors but will rarely be relevant to their interests. Although there has been considerable progress in recent years in comparative research addressing the capacities of nonhuman primates and birds to model the minds of others (Bugnyar, Stöwe, & Heinrich, 2004; Hare, Call, Agnetta, & Tomasello, 2000; Hare & Tomasello, 2004), the high-level ToM involved in human language appears to be another "key innovation" our species acquired before language could reach its full, modern form. Because the possession of language itself almost certainly aids our capacity to model the minds of others, this may be a chicken-and-egg question: without ToM you can not have language, and without language you do not have full ToM. Resolution of this issue remains a central issue in the evolution of language.

Syntax Syntax is a core component of language because of its mediating role between the perceptual world of speech and sign and the conceptual world of thoughts, imagery, and memory. Because it is clear that our prelinguistic ancestors had both complex perceptual and motor control, and complex cognition, syntax is often supposed to be the most recent evolutionary advance, and thus to constitute a "key innovation" during the phylogenesis of human language. Nonetheless, there is every reason to expect that syntax built upon a preexisting ability of organisms to find patterns in perceptual input, and that these pattern-discovery mechanisms can be studied comparatively.

One promising development in this regard is the use of "artificial grammar learning" to probe spontaneous capacities for pattern induction. Such work involves exposing subjects to a set of experimentally generated stimuli (the "exposure" phase) and then testing to see what patterns participants have extracted, by examining their reactions to novel stimuli that either fit into or violate the "rules" implicit in the exposure set. This work was originally developed with adult humans (Reber, 1967), but methodological advances now allow similar techniques to be used with human infants (Gómez & Gerken, 1999; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Aslin, & Newport, 1996) and nonhuman animals (Hauser, Newport, & Aslin, 2001; Hauser, Weiss, & Marcus, 2002; Toro & Trobalón, 2005). Using the well-developed formalism of formal language theory to generate stimuli, Fitch and Hauser (2004) tested the abilities of humans and cotton-top tamarin monkeys to learn either a simple finite-state grammar or a more complex phrase-structure grammar. The results suggest that these monkeys are limited to the former class. These results have been extended with other species, and these recent data suggest that the ability to parse hierarchical phrase structure relies on a particular computational component that is clearly present in humans, lacking in at least some nonhuman primates, and perhaps present in starlings (a songbird species) (Fitch & Hauser, 2004; Gentner, Fenn, Margoliash, & Nusbaum, 2006). Furthermore, brain-imaging data suggest that humans process such grammar classes using different cytoarchitectonic regions in the inferior frontal cortex (Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006). Although the genetic mechanisms involved remain unstudied at present, this may be another example where birds provide a better model species for human language than nonhuman primates. If so, this will be excellent news for neurobiologists interested in understanding such mechanisms, since empirical techniques applicable to the developing avian brain are, and seem likely to remain, more sophisticated and advanced than those available for nonhuman primates (and of course for humans).

Conclusion

While the genetic bases underlying syntactic or semantic abilities in our species remain unknown, there is good reason to expect rapid progress in uncovering them. This will usher

in an exciting new world, both for cognitive neuroscientists interested in tying complex cognition to the underlying neural architecture and for evolutionary biologists interested in uncovering the phylogenetic trajectory that led to human language. In this new era, the identification of deep homologies may play a central role. This is excellent news for comparative biologists, because it suggests that a far broader range of vertebrates, and even nonchordates, may offer valuable windows into the genetic basis of that most human of traits, language.

REFERENCES

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.

Barkow, J., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind*. Oxford, UK: Oxford University Press.

Bolhuis, J. J., & Macphail, E. M. (2001). A critique of the neuroecology of learning and memory. *Trends Cogn. Sci.*, *5*(10), 426–433.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication*. Sunderland, MA: Sinauer Associates.

Bugnyar, T., Stöwe, M., & Heinrich, B. (2004). Ravens, *Corvus corax*, follow gaze direction of humans around obstacles. *Proc. R. Soc. Lond. B Biol. Sci.*, *271*(1546), 1331–1336.

Call, J., & Tomasello, M. (2007). *The gestural communication of apes and monkeys*. London: Lawrence Erlbaum.

Camhi, J. M. (1984). *Neuroethology: Nerve cells and the natural behavior of animals*. Sunderland, MA: Sinauer Associates.

Carroll, S. B. (2003). Genetics and the making of *Homo sapiens*. *Nature*, *422*(6934), 849–857.

Carroll, S. B. (2005). *Endless forms most beautiful*. New York: W. W. Norton.

Carroll, S. B. (2006). *The making of the fittest: DNA and the ultimate forensic record of evolution*. New York: W. W. Norton.

Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2005). *From DNA to diversity: Molecular genetics and the evolution of animal design* (2nd ed.). Malden, MA: Blackwell.

Chen, F.-C., & Li, W.-H. (2001). Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.*, *68*(2), 444–456.

Chimpanzee Sequencing and Analysis Consortium. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, *437*, 69–87.

Christiansen, M., & Kirby, S. (Eds.). (2003). *Language evolution*. Oxford, UK: Oxford University Press.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: John Murray.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology* (pp. 282–309). Oxford, UK: Blackwell.

De Robertis, E. M., & Sasai, Y. (1996). A common plan for dorsoventral patterning in Bilateria. *Nature*, *380*, 37–40.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: W. W. Norton.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annu. Rev. Neurosci.*, *22*, 567–631.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: A basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development* (Suppl.), *1994*, 135–142.

Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development*, *134*(14), 2549–2560.

Enard, W., Khaitovich, P., Klose, J., & Pääbo, S. (2002). Intra- and interspecific variation in primate gene expression patterns. *Science*, *296*, 340–343.

Fay, J. C., Wyckoff, G. J., & Wu, C.-I. (2001). Postive and negative selection on the human genome. *Genetics*, *158*, 1227–1234.

Fernald, R. D. (2000). Evolution of eyes. *Curr. Opin. Neurobiol.*, *10*, 444–450.

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nat. Genet.*, *18*(2), 168–170.

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends Cogn. Sci.*, *4*(7), 258–267.

Fitch, W. T. (2004). Kin selection and "mother tongues": A neglected component in language evolution. In D. K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach* (pp. 275–296). Cambridge, MA: MIT Press.

Fitch, W. T. (2007a). Evolving meaning: The roles of kin selection, allomothering and paternal care in language evolution. In C. Lyon, C. Nehaniv, & A. Cangelosi (Eds.), *Emergence of communication and language* (pp. 29–51). New York: Springer.

Fitch, W. T. (2007b). Linguistics: An invisible hand. *Nature*, *449*, 665–667.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*, 377–380.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, *97*(2), 179–210.

Foster, K. R., Wenseleers, T., & Ratnieks, F. L. W. (2006). Kin selection is the key to altruism. *Trends Ecol. Evol.*, *21*(2), 57–60.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proc. Natl. Acad. Sci. USA*, *103*(7), 2458–2463.

Gehring, W. J., & Ikeo, K. (1999). Pax 6: Mastering eye morphogenesis and eye evolution. *Trends Genet.*, *15*, 371–377.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*, 1204–1207.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.

Grice, H. P. (1975). Logic and conversation. In D. Davidson & G. Harman (Eds.), *The logic of grammar* (pp. 64–153). Encino, CA: Dickenson.

Groszer, M., Keays, D., Deacon, R., de Bono, J., Prasad-Mulcare, S., Gaub, S., et al. (2008). Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr. Biol.*, *18*, 354–362.

Haesler, S., Rochefort, C., Geogi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol*, *5*, e321.

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., et al. (2004). FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.*, *24*, 3164 –3175.

Hamilton, W. D. (1964). The evolution of altruistic behavior. *Am. Nat.*, *97*, 354–356.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Anim. Behav.*, *59*(4), 771–785.

Hare, B., & Tomasello, M. (2004). Chimpanzees are more skillful in competitive than cooperative cognitive tasks. *Anim. Behav.*, *68*, 571–581.

Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The language faculty: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.

Hauser, M. D. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*, 53–64.

Hauser, M. D., Weiss, D., & Marcus, G. (2002). Rule learning by cotton-top tamarins. *Cognition*, *86*, B15–B22.

Hayes, C. (1951). *The ape in our house*. New York: Harper.

Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *J. Comp. Physiol. Psychol.*, *56*, 872–876.

Heyes, C. (2003). Four routes of cognitive evolution. *Psychol. Rev.*, *110*(4), 713–727.

Hubel, D., & Wiesel, T. (1965). Binocular interaction in striate cortex of kittens reared with aritificial squint. *J. Neurophysiol.*, *28*, 1041–1059.

Hull, R., & Vaid, J. (2007). Bilingual language lateralization: A meta-analytic tale of two hemispheres. *Neuropsychologia*, *45*, 1987–2008.

Hurford, J. (1990). Nativist and functional explanations in language acquisition. In I. M. Roca (Ed.), *Logical issues in language acquistion* (pp. 85–136). Dordrecht: Foris.

Janik, V. M., & Slater, P. B. (1997). Vocal learning in mammals. *Adv. Study Behav.*, *26*, 59–99.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Ann. NY Acad. Sci.*, *1016*, 749–777.

Jarvis, E. D. (2007). Neural systems for vocal learning in birds and humans: A synopsis. *J. Ornithol.*, *143*, S35–44.

Jürgens, U. (1994). The role of the periaqueductal grey in vocal behaviour. *Behav. Brain Res.*, *62*, 107–117.

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.*, *26*(2), 235–258.

Kako, E. (1999). Elements of syntax in the systems of three language-trained animals. *Anim. Learn. Behav.*, *27*, 1–14.

Kirby, S., Smith, K., & Brighton, H. (2004). From UG to universals: Linguistic adaptation through iterated learning. *Stud. Lang.*, *28*(3), 587–607.

Kuypers, H. G. J. M. (1958a). Corticobulbar connections to the pons and lower brainstem in man: An anatomical study. *Brain*, *81*, 364–388.

Kuypers, H. G. J. M. (1958b). Some projections from the pericentral cortex to the pons and lower brain stem in monkey and chimpanzee. *J. Comp. Neurol.*, *110*, 221–255.

Laland, K. N., & Brown, G. R. (2002). *Sense and nonsense: Evolutionary perspectives on human behaviour*. Oxford, UK: Oxford University Press.

Larson, C. R., Ortega, J. D., & DeRosier, E. A. (1988). Studies on the relation of the midbrain periaqueductal gray, the larynx and vocalization in the awake monkey. In J. D. Newman (Ed.), *The physiological control of mammalian vocalizations* (pp. 43–65). New York: Plenum Press.

Larson, C. R., Sutton, D., Taylor, E. M., & Lindeman, R. (1973). Sound spectral properties of conditioned vocalizations in monkeys. *Phonetica*, *27*, 100–112.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature*, *276*, 565–570.

Lieberman, P. (1998). *Eve spoke: Human language and human evolution*. New York: W. W. Norton.

Lieberman, P. (2000). *Human language and our reptilian brain: The subcortical bases of speech, syntax and thought*. Cambridge, MA: Harvard University Press.

Liégeois, F., Connelly, A., Cross, J., Boyd, S. G., Gadian, D. G., Vargha-Khadem, F., et al. (2004). Language reorganization in children with early-onset lesions of the left hemisphere: An fMRI study. *Brain*, *127*(6), 1229–1236.

Liem, K. (1973). Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws. *Syst. Zool.*, *22*, 425–441.

Locke, J., & Pearson, D. M. (1990). Linguistic significance of babbling: Evidence from a tracheostomized infant. *J. Child Lang.*, *17*, 1–16.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80.

Marler, P., & Slabbekoorn, H. (2004). *Nature's music: The science of birdsong*. New York: Academic Press.

Maynard Smith, J. (1964). Group selection and kin selection. *Nature*, *201*, 1145–1147.

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford, UK: Oxford University Press.

Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Harvard University Press.

Nettle, D. (1999). *Linguistic diversity*. Oxford, UK: Oxford University Press.

Nottebohm, F. (1989). From bird song to neurogenesis. *Sci. Am.*, *260*, 74–79.

Nottebohm, F., Kasparian, S., & Pandazis, C. (1981). Brain space for a learned task. *Brain Res.*, *213*, 99–109.

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*, 717–721.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*(2), 201–236.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature*, *434*, 455–456.

Premack, D. (1986). *Gavagai! Or the future history of the animal language controversy*. Cambridge, MA: MIT Press.

Rakic, P. (1985). Limits of neurogenesis in primates. *Science*, *227*, 1054–1055.

Reber, A. S. (1967). Implicit learning of artificial grammars. *J. Verb. Learn. Verb. Beh.*, *6*(6), 855–863.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

Rutishauser, R., & Moline, P. (2005). Evo-devo and the search for homology ("sameness") in biological systems. *Theory Biosci.*, *124*, 213–241.

Saffran, J., Aslin, D., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Scharff, C., & Haesler, S. (2005). An evolutionary perspective on FoxP2: Strictly for the birds? *Curr. Opin. Neurobiol.*, *15*, 694–703.

Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*, 699–701.

Schnitzler, H.-U., Menne, D., Kober, R., & Heblich, K. (1983). The acoustical image of fluttering insects in echolocating bats. In F. Huber & H. Markl (Eds.), *Neuroethology and behavioral physiology: Roots and growing pains* (pp. 235–251). Berlin: Springer-Verlag.

Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., et al. (2005). Altered ultrasonic vocalization in mice with a disruption in the *Foxp2* gene. *Proc. Natl. Acad. Sci. USA*, *102*(27), 9643–9648.

Shubin, N. (2008). *Your inner fish: A journey into the 3.5 billion-year history of the human body*. London: Penguin Books.

Shubin, N., Tabin, C., & Carroll, S. (1997). Fossils, genes and the evolution of animal limbs. *Nature*, *388*, 639–648.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford, UK: Blackwell.

Striedter, G. F. (2004). *Principles of brain evolution*. Sunderland, MA: Sinauer.

Tomarev, S. I., Callaerts, P., Kos, L., Zinovieva, R., Halder, G., Gehring, W., et al. (1997). Squid Pax-6 and eye development. *Proc. Natl. Acad. Sci. USA*, *94*(6), 2421–2426.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2001). Cultural transmission: A view from chimpanzees and human infants. *J. Cross Cult. Psychol.*, *32*(2), 135–146.

Tomasello, M. (2007). If they're so good at grammar, then why don't they talk? Hints from apes' and humans' use of gestures. *Lang. Learn. Dev.*, *3*(2), 133–156.

Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford, UK: Oxford University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behav. Brain Sci.*, *28*, 675–735.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Percept. Psychophys.*, *67*(5), 867–875.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.*, *46*, 35–57.

van Heyningen, V., & Williamson, K. A. (2002). PAX6 in sensory development. *Hum. Mol. Genet.*, *11*(10), 1161–1167.

Vargha-Khadem, F., Gadian, D. G., Copp, A., & Mishkin, M. (2005). FOXP2 and the neuroanatomy of speech and language. *Nat. Rev. Neurosci.*, *6*(2), 131–138.

Vargha-Khadem, F., & Passingham, R. (1990). Speech and language deficits. *Nature*, *346*, 226.

Vargha-Khadem, F., Watkins, K., Price, C. J., Ashburner, J., Alcock, K., Connelly, A., et al. (1998). Neural basis of an inherited speech and language disorder. *Proc. Natl. Acad. Sci. USA*, *95*, 12695–12700.

von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, *404*(6780), 871–876.

Wada, K., Howard, J. T., McConnell, P., Whitney, O., Lints, T., Rivas, M., et al. (2006). A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes. *Proc. Natl. Acad. Sci. USA*, *103*(41), 15212–15217.

Wild, J. M. (1993). The avian nucleus retroambigualis: A nucleus for breathing, singing and calling. *Brain Res.*, *606*, 119–124.

Wilkins, A. S. (2002). *The evolution of developmental pathways*. Sunderland, MA: Sinauer.

Zahavi, A. (1993). The fallacy of conventional signalling. *Proc. R. Soc. Lond. B Biol. Sci.*, *340*, 227–230.

VIII
THE EMOTIONAL AND SOCIAL BRAIN

Chapter 60 SULLIVAN, MORICEAU, RAINEKI, AND ROTH 889

61 LEDOUX, SCHILLER, AND CAIN 905

62 VUILLEUMIER AND BROSCH 925

63 WHALEN AND DAVIS 935

64 HARIRI 945

65 MITCHELL AND HEATHERTON 953

66 BEER 961

67 SINGER AND LEIBERG 973

68 GREENE 987

Introduction

TODD F. HEATHERTON AND
JOSEPH E. LEDOUX

IN THE PREVIOUS edition of this book we noted that the neuroscientific study of emotion was flourishing, and this led scientists to recognize that other hot aspects of cognition—namely, the social aspects of cognition—were equally due for attention. In the past five years the field of social neuroscience has exploded, as evidenced by the launching of new journals and major initiatives from many federal funding agencies. At the core of this emphasis on social cognition is the importance of emotional processing. The social and emotional aspects of the brain are inexorably linked, the adaptive significance of emotions being closely linked to their social value, and nearly all social interaction produces affective responses. It is clear from the chapters in this section that the study of emotion and social cognition remains at the forefront of cognitive neuroscience.

The interdisciplinary field of cognitive neuroscience has provided ample evidence of the benefits of examining psychological constructs across multiple levels of analysis, from the molecular to the cultural. The chapters in this section cross many levels, from the role of systems, cellular and molecular mechanisms in learning and memory in adult animals (LeDoux, Schiller, and Cain) and during early development (Sullivan and colleagues), to human studies of emotion (Vuilleumier and Brosch, Whalen and Davis) and emotion regulation (Beer) using functional imaging, including imaging genomics (Hariri), and studies of brain regions that support social cognition (Mitchell and Heatherton), sensitivity to others (Singer and Leiberg), and moral transgressions (Greene), many of the chapters emphasizing how these systems interact. It is clear that research that uses multiple approaches is moving the field toward the ultimate goal of developing coherent models of how the brain makes emotion and performs its social function.

In the last edition there was remarkable convergence of evidence highlighting the important role of the amygdala across animal species and paradigms. The importance of the amygdala continues to hold center stage in the neuroscience of emotion, as is evident in many of the chapters. LeDoux, Schiller, and Cain provide a timely review of the importance of the amygdala in normal and pathological fear, especially the cellular molecular mechanisms that underlie learning and memory. For instance, they update research demonstrating the important role of gene expression in long-term potentiation, which is perhaps the most likely candidate for the physiological basis of fear conditioning and other forms of memory. The chapter by Sullivan and colleagues demonstrates that the role of the amygdala in fear learning depends on its developmental maturity. During the sensitive period associated with a rat pup's attachment to its mother (about 10 days postnatal), fear learning is attenuated due to a lack of amygdala functioning. This process allows the pup to attach to its mother irrespective of whether the mother is treating it harshly or gently. Vuilleumier and Brosch present exciting new evidence that attention and emotion can influence basic visual processing. That is, the emotional significance of stimuli, not just the strictly visual properties of the retinal image, influences basic perceptual processing. They provide compelling evidence that the amygdala plays an important role in the modulation of sensory information. Whalen and Davis address how context influences amygdala activity and subsequent interpretations of emotional stimuli; that is, they describe how amygdala activity is strongly affected by the context in which biologically relevant cues are encountered. Hariri provides examples of how imaging genetics can lead to insights about the biological mechanisms underlying individual differences in complex behavioral traits, such as how abnormal gene expression in the serotonin system can affect how the amygdala responds to facial expressions. Taken together, these chapters not only reinforce the important role of the amygdala in emotional processing but also demonstrate how new techniques and approaches are continuing to provide important insights.

The last four chapters in this section focus more on the social basis of emotional processing. These chapters demonstrate that there continues to be growing interest among social psychologists and cognitive neuroscientists in using brain imaging to study social aspects of cognition, such as recognition of faces and emotional expressions, theory of mind, social emotions such as empathy, judging trustworthiness and attractiveness, and cooperation. Mitchell and Heatherton outline the basic components necessary for a social brain, including a sense of self, mentalizing ability, capacity for self-regulation, and threat detectors for ingroup and outgroup threats, and they discuss the discrete neural signatures of these basic components. Beer describes research on how people are able to regulate their emotions so that they can function in their social groups. She outlines evidence that damage to areas of frontal cortex interferes with emotion regulation. Singer and Leiberg describe fascinating new research on the neural basis of empathy and how it is influenced by both dispositional and contextual factors. Finally, Greene describes research on how human brains make moral judgments. The ethical brain reflects perhaps the greatest achievement of human evolution, and researchers are just beginning to identify how people make these types of important decisions.

Considering the section as a whole, it is clear that the study of emotion continues to be a strong growth area in cognitive neuroscience. Moreover, it has expanded to include the closely connected social brain. As an organ that has evolved to solve adaptive problems, the brain relies on emotional processes to solve challenges to successful adaptation. For humans, many of the most pernicious adaptive problems involve other humans, such as selecting mates, cooperating in hunting and gathering, forming alliances, competing over scarce resources, and even warring with neighboring groups. Interacting with other humans produces emotion, and these emotions serve as guidelines for successful group living. For example, behaviors such as lying, cheating, and stealing are discouraged by social norms in all societies because they decrease survival and reproduction for other group members. They also elicit vigorous emotional responses. Hence any true understanding of human nature will require a full consideration of both the emotional brain and the social brain. We expect that research on this topic will continue to be on the cutting edge of cognitive neuroscience in the next decade.

60 Ontogeny of Infant Fear Learning and the Amygdala

REGINA M. SULLIVAN, STEPHANIE MORICEAU, CHARLIS RAINEKI, AND TANIA L. ROTH

ABSTRACT To support attachment to the caregiver, altricial infants, including humans and rats, must identify, learn, and remember their caregiver. The early attachment process in the rat is distinguished by its behavior and underlying neural circuitry, which are both exquisitely suited to promoting the infant-caregiver relationship. Foremost, infants have the enhanced ability to acquire learned preferences, and this behavior is supported by the hyperfunctioning locus coeruleus and experience-induced changes in the olfactory bulb and anterior piriform cortex. But of equal importance, infants have a decreased ability to acquire learned aversions and fear, and this behavior is facilitated through attenuated amygdala activity. Presumably, this attachment circuitry constrains the infant to form only preferences for the caretaker regardless of the quality of the care received. With maturation and the end of the infant-caregiver attachment learning period, the developing rat's social behavior and underlying circuitry transition to accommodate life outside the nest. However, early-life environmental and physiological stressors can alter the dynamic nature of this circuitry, particularly in respect to the amygdala. Such changes likely provide a framework for the lasting effects of early stress on emotional and cognitive outcome.

The altricial infant's social world revolves around the caregiver, and as evolution would have it, the infant's emotional and social behaviors have been well crafted to form and maintain the infant-caregiver relationship. Infants of many altricial species must learn to recognize their caregiver as the target of their social behavior and continue to express proximity-seeking behaviors toward their caregiver to receive the food, protection, and warmth necessary for survival. This learning about the caregiver and the emergence of social behavior directed toward the caregiver are referred to as attachment, and this process has wide phylogenetic representation, including chicks, rodents, nonhuman primates, and humans.

REGINA M. SULLIVAN, STEPHANIE MORICEAU, and CHARLIS RAINEKI Emotional Brain Institute, Nathan Kline Institute and Child and Adolescent Psychiatry, New York University Langone Medical Center, Orangeburg, New York
TANIA L. ROTH Department of Neurobiology and the Evelyn F. McKnight Brain Institute, University of Alabama at Birmingham, Birmingham, Alabama

The social environment of the developing altricial animal is very different at birth and weaning. For example, social behavior in the infant rat following birth is limited to proximity-seeking of the caregiver. And though the complex social behavior of the developing and preweanling rat pup still involves proximity-seeking of the caretaker, it now must also facilitate interactions with peers as well as the unfamiliar social world outside the nest. Thus the rapid maturation of most altricial mammals and the ultimate transition to adult social behavior require dynamic neural circuitry that is capable of responding to these contrasting environments. In this chapter, we will review the literature on infant attachment learning and the underlying neural circuitry that mediate early infant-caregiver social interactions, the transitioning role of this behavior and circuitry during development, and the enduring effects of stress on both the attachment circuitry and adult behavior.

Early-life social behavior: Attachment learning

Altricial infants of many species, including the human and rat, must learn to identify, orient and approach, and prefer their own mother (Bowlby, 1969; Polan & Hofer, 1998; Shair, Masmela, Brunelli, & Hofer, 1997). This attachment learning begins during fetal life and continues after birth. For example, human infants recognize, orient toward, and prefer their own mother's voice when tested within hours of birth (DeCasper & Fifer, 1980). Furthermore, two-day-old newborns will increase suckling at the sound of their own mother's voice versus any other human voice, indicative of a learned preference for maternal voice (Fifer & Moon, 1995). This recognition is also true regarding maternal odor. At birth, a human infant who is placed on the mother's ventrum will slowly approach a breast scented with amniotic fluid in preference to an untreated breast (Varendi, Porter, & Winberg, 1996), and a change in maternal diet, which will alter the odor of the amniotic fluid, directly influences this preferential response (Lecanuet & Schaal, 1996; Mennella, Johnson, & Beauchamp, 1995; Schaal, Marlier, & Soussignan, 1995). This early odor preference appears to be learned and modulates interaction with

the mother (Schaal, Marlier, & Soussignan, 1995; Sullivan et al., 1991).

Odor learning about the mother for infant attachment appears phylogenetically widespread. Similar learning controls early social behavior in rats (Alberts & May, 1984; Blass & Teicher, 1980; Polan & Hofer, 1998; Risser & Slotnick, 1987; Teicher, Flaum, Williams, Eckhert, & Lumia, 1978), rabbits (Distel & Hudson, 1985; Hudson, 1985; Hudson & Distel, 1983), and mice (Armstrong, DeVito, & Cleland, 2006; Coppola, Coltrane, & Arsov, 1994; M. B. Hennessy, Li, & Levine, 1980; Moles, Kieffer, & D'Amato, 2004). In these species, an infant's social world after birth is the nest; therefore social behavior is mostly directed toward the mother. Indeed, during this time of dependency upon the mother, behavior is centered on maintaining contact with the mother, and this behavior is guided and controlled by the presence of maternal odor (Galef & Kaner, 1980; Leon, 1992). Specifically, maternal odor drives an infant to approach the mother and induces nipple attachment, while chemical removal of the natural maternal odor disrupts these behaviors (Hofer, Shair, & Singh, 1976; Teicher & Blass, 1977).

Infant rats learn their mother's odor naturally within the nest (Brunjes & Alberts, 1979; Campbell, 1984; Galef & Kaner, 1980; Leon, 1975; Miller, Jagielo, & Spear, 1989; Pedersen, Williams, & Blass, 1982; Rudy & Cheatle, 1977; Sullivan, Brake, Hofer, & Williams, 1986; Sullivan, Hofer, & Brake, 1986; Sullivan, Wilson, Wong, Correa, & Leon, 1990; Terry & Johanson, 1996). However, this learning can be mimicked in classical conditioning experiments outside the nest (Camp & Rudy, 1988; Haroutunian & Campbell, 1979; Moriceau & Sullivan, 2006; Roth & Sullivan, 2005; Spear, 1978; Sullivan, Hofer, & Brake, 1986; Sullivan, Landers, Yeaman, & Wilson, 2000). Specifically, paired presentations of odor and reward are sufficient to produce both learned odor preferences (demonstrated by an approach to the odor) and nipple attachment. Furthermore, as illustrated in figure 60.1, a broad range of stimuli have been shown to function as a reward capable of producing learned odor preferences in infant rats outside the nest (Alberts & May, 1984; Brake, 1981; Galef & Sherry, 1973; Johanson & Hall, 1979; Johanson & Teicher, 1980; Leon, 1975; McLean, Darby-King, Sullivan, & King, 1993; Pedersen et al., 1982; Sullivan, Brake, et al., 1986; Sullivan, Hofer, & Brake, 1986; Weldon, Travis, & Kennedy, 1991; Wilson & Sullivan, 1994).

Though it is well established that the infant rat shows excellent learning and memory ability, particularly for learned odor preferences, we still understand very little of the neural framework that is responsible for this early behavior. Indeed, the neural structures that are well documented to support learned behavior in adult rats (e.g., hippocampus, frontal cortex, and amygdala) are not yet fully functional in infants. This suggests that the neural circuitry that mediates attachment learning and memory in the developing rat might differ from that in the adult. Our work as well as that of others has shown that indeed this is the case. Together, data implicate a unique neural framework in the infant that is responsible for the olfactory-based attachment learning.

Attachment learning circuitry

Both anatomical and physiological changes within the olfactory bulb have been documented to support odor preference learning and memory in infant rats (Fillion & Blass, 1986; Fleming, O'Day, & Kraemer, 1999; Johnson, Woo, Duong, Nguyen, & Leon, 1995; Moore, Jordan, & Wong, 1996; Sullivan & Wilson, 1991; Wilson, Sullivan, & Leon, 1987; Woo, Coopersmith, & Leon, 1987; Yuan, Harley, Darby-King, Neve, & McLean, 2003; Zhang, Okutani, Inoure, & Kaba, 2003). These changes occur not only in response to odors experienced in the nest (Sullivan et al., 1990), but also in controlled learning experiments outside the nest (Sullivan & Leon, 1986; Johnson et al., 1995; Moriceau & Sullivan,

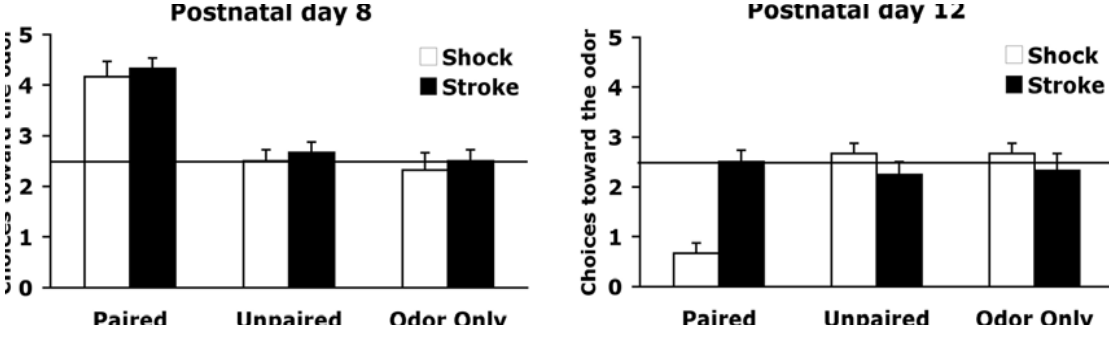


FIGURE 60.1 This graph illustrates pup preference learning from stroking (mimicking mother licking) and shock (mimicking pain received from mother) and developmental changes. During a sensitive period, eight-day-old rat pups readily form a learned odor preference to contiguous presentations of odor and stroking or tail- or foot-shock (0.5 mA). With the close of the sensitive period, twelve-day-old pups no longer show learned-odor associations with stroking and, in contrast to younger pups, show learned aversions to odor-shock presentations.

2004; Roth & Sullivan, 2005; Yuan, Harley, McLean, & Knopfel, 2002; Sullivan & Wilson, 1991; Wilson, Sullivan, & Leon, 1987). These learning-induced olfactory bulb changes are attributable to the large influx of norepinephrine (NE) released from the locus coeruleus (LC) (McLean & Shipley 1991; Shipley, Halloran, & de la Torre, 1985), which prevents the mitral cells of the olfactory bulb from habituating to continual olfactory stimulation (Okutani, Kaba, Takahashi, & Seto, 1998; Sullivan, Stackenwalt, Nasr, Lemon, & Wilson, 2000; Sullivan, Wilson, & Leon, 1989; Sullivan, Zyzak, Skierkowski, & Wilson, 1992; Wilson, Sullivan, & Leon, 1987). At the molecular level, NE also increases CREB phosphorylation (pCREB) via cAMP stimulation (McLean, Harley, Darby-King, & Yuan, 1999; Yuan et al., 2003; Zhang et al., 2003). This ultimately activates the transcription of immediate-early and late-response genes whose readouts support synapse formation, neurogenesis, and learning (Bekinschtein, Cammarota, Izquierdo, & Medina, 2008; Tao, Finkbeiner, Arnold, Shaywitz, & Greenberg, 1998). This is a common cellular cascade that mediates learned behavior in many species across development (Carew, 1996; Carew & Sutton, 2001; Rankin, 2002).

In the infant, the abundant amount of NE released to the olfactory bulb is inducible by a wide range of sensory stimuli (Nakamura & Sakaguchi, 1990; Rangel & Leon, 1995). Furthermore, NE is both necessary and sufficient for the learning-induced behavioral and neural changes that are displayed in the infant. Specifically, either lesioning the LC or blocking NE receptors in the olfactory bulb prevents associative learning. An odor preference is also readily learned with paired presentations of NE (LC stimulation or intrabulbar NE infusions) and an odor (Sullivan et al., 1992; Sullivan, Wilson, Lemon, & Gerhardt, 1994; Sullivan, Stackenwalt, 2000; Yuan et al., 2003; Langdon, Harley, & McLean, 1997).

The infant LC releases substantially more NE into the olfactory bulb than the adult LC (Rangel & Leon, 1995). In fact, there is a sharp contrast between the functioning of the infant and adult LC. Unlike the adult LC, the infant LC responds to a broader range of sensory stimuli and fails to habituate after repeated sensory stimulation (Kimura & Nakamura, 1985; Nakamura & Sakaguchi, 1990; Foote, Aston-Jones, & Bloom, 1980; Harley & Sara, 1992; Sara, Dyon-Laurent, & Herve, 1995; Vankov, Herve-Minvielle, & Sara, 1995). Differences in the functioning of the autoreceptors that are located on the somadendritic membranes of the LC neurons appear to be responsible for the development-related differences in stimulus-response times (Marshall, Christie, Finlayson, & Williams, 1991; Nakamura & Sakaguchi, 1990; Winzer-Serhan, Raymon, Broide, Chen, & Leslie, 1997). Specifically, an infant LC's $\alpha 2$ inhibitory autoreceptors, while present, do not appear functional; therefore the excitatory $\alpha 1$ autoreceptors ensure a prolonged response upon sensory stimulation. At the end of the attachment learning period, "adultlike" characteristics of the LC emerge, including a shorter stimulus-evoked response time and stimulus habituation. These adult characteristics parallel the functional emergence of the $\alpha 2$ inhibitory autoreceptors, and this functional change in physiology of the LC dovetails when NE begins to play a more modulatory role (verses the necessary and sufficient role) in learned behavior (Ferry & McGaugh, 2000; McGaugh, 2006).

Thus far, we have discussed the experience-induced changes in the olfactory bulb and LC that are responsible for early-life learned-odor associations. One additional structure that we recently added to the infant attachment circuit is the piriform "olfactory" cortex. Axons of the mitral cells of the olfactory bulb project directly to the piriform cortex (Haberly, 2001; Schwob & Price, 1984a), and the piriform cortex can be divided, both anatomically and physiologically, into two distinct structures: (1) the anterior piriform cortex, which is more influenced by direct olfactory bulb input, and (2) the posterior piriform cortex, which is more influenced by input from other limbic structures and intracortical connectivity (Swanson & Petrovich, 1998; Wilson & Stevenson, 2003). While cellular and physiological changes in the olfactory bulb play a prominent role in the early learning process, the piriform cortex appears to have an important role in assigning the hedonic value to a learned odor. In particular, early-life learned odor preferences engage the anterior piriform cortex (with no detectable activity in the posterior piriform), while learned odor aversions in older pups and adults engage posterior piriform cortical activity (Moriceau & Sullivan 2006; Moriceau et al., 2006; Raineki, Shionoya, Sander, & Sullivan, 2009; Roth & Sullivan 2005).

In summary, the data that we have reviewed thus far suggest that the contingent events of stimulus-induced NE release from the LC and NE-induced physiological and molecular changes in the olfactory bulb and anterior piriform cortex support the neural plasticity that is responsible for the acquisition of olfactory-based attachment behavior in the infant rat.

Fear and amygdala are attenuated in early life

In addition to enhanced preference learning supported by the neural circuitry discussed above, infant social behavior is also characterized by limitations on aversive learning. For example, shocking a chick during imprinting actually enhances following of the surrogate caregiver, although shock supports avoidance just hours after the imprinting critical period closes (Hess, 1962; Rajecki, Lamb, & Obmascher, 1978; Salzen, 1970). Similarly, shocking an infant dog or rat results in a strong attachment to the caregiver (Camp & Rudy, 1988; Roth & Sullivan, 2005; Spear, 1978; Stanley, 1962; Sullivan, Brake, et al., 1986;

Sullivan, Landers, et al., 2000). Finally, nonhuman primate and human infants exhibit strong proximity-seeking behavior toward an abusive mother (Harlow & Harlow, 1965; Maestripieri, Tomaszycki, & Carroll, 1999; Sanchez, Ladd, & Plotsky, 2001; Suomi, 2003).

Certain types of inhibitory learning, including fear of predators, cued- and contextual-fear conditioning, inhibitory conditioning, and passive avoidance, do not emerge until after postnatal days 10–11 (Blozovski & Cudennec, 1980; Camp & Rudy, 1988; Collier, Mast, Meyer, & Jacobs, 1979; Goldman & Tobach, 1967; Haroutunian & Campbell, 1979; Myslivecek, 1997; Stehouwer & Campbell, 1978; Sullivan, Landers, et al., 2000). Indeed, aversive stimuli such as moderate shock (as shown in figure 60.1) and tailpinch elicit learned odor preferences in infant rats (Camp & Rudy, 1988; Haroutunian & Campbell, 1979; Moriceau & Sullivan, 2006; Moriceau et al., 2006; Roth & Sullivan, 2005; Spear, 1978; Sullivan, Hofer, and Brake, 1986; Sullivan, Landers, et al., 2000), despite an apparent pain response (Barr, 1995; Collier & Bolles, 1980; Emerich, Scalzo, Enters, Spear, & Spear, 1985; Fitzgerald, 2005; Shair, Masmela, Brunelli, & Hofer, 1997; Stehouwer & Campbell, 1978).

What could explain this paradoxical preference learning to aversive stimuli? Evidence suggests that the lack of amygdala plasticity may play a leading role. Indeed, the limitations on fear learning, passive avoidance, active avoidance, and inhibitory conditioning during the sensitive period correspond to the period during development when the amygdala does not participate in the learning process (Blozovski & Cudennec, 1980; Collier et al., 1979; Myslivecek, 1997). Specifically, the amygdala is not evoked during infant learning in classical fear-conditioning or natural fear paradigms (Moriceau, Roth, Okotoghaide, & Sullivan, 2004; Moriceau & Sullivan, 2006; Moriceau et al., 2006; Roth & Sullivan, 2005; Wiedenmayer & Barr, 2001). On the contrary, in other animals ranging from *Caenorhabditis-elegans* to rodents and humans, the amygdala is a brain area that is readily evoked by aversive stimuli in classical conditioning and natural fear paradigms (Blair, Schafe, Bauer, Rodrigues, & LeDoux, 2001; Davis, 1997; Fanselow & Gale, 2003; Fanselow & LeDoux, 1999; Herzog & Otto, 1997; Maren, 2003; McGaugh, Roozendaal, & Cahill, 1999; Pape & Stork, 2003; Pare, Quirk, & Ledoux, 2004; Rosenkranz & Grace, 2002; Sananes & Campbell, 1989; Schettino & Otto, 2001; Sevelinges, Gervais, Messaoudi, Granjon, & Mouly, 2004; Sigurdsson, Doyere, Cain, & LeDoux, 2007).

One contributing factor for the apparent lack of amygdala plasticity in early-life learning may be functional amygdala immaturity. The development of the amygdala is considered protracted and extends into adolescence, though peak neurogenesis and nuclei subdivision occur as early as the first week of life (Bayer, 1980; Berdel & Morys, 2000a, 2000b; Berdel, Morys, & Maciejewska, 1997; Bouwmeester, Smits, & Van Ree, 2002; Cunningham, Bhattacharyya, & Benes, 2002; Morys, Berdel, Jagalska-Majewska, & Luczynska, 1999; Nair & Gonzalez-Lima, 1999). Synaptic development begins to appear around PN5, with a dramatic increase between PN10 and PN20, but adult levels are not reached until early adolescence (Mizukawa, Tseng, & Otsuka, 1989). Furthermore, the typical long-term synaptic plasticity (LTP) that is inducible in the adult basolateral amygdala does not emerge until the end of the attachment learning period (Thompson, Sullivan, & Wilson, 2008).

Thus far in this review, we have discussed the literature that presents the case that it is difficult for infants to learn aversions. Sadly, attachment occurs regardless of inadequate caregiving. Specifically, children tolerate considerable abuse while remaining strongly attached to an abusive caretaker (Helfer, Kempe, & Krugman, 1997; Pollak, 2003). Moreover, attachment despite abuse is spread throughout the animal kingdom (Camp & Rudy, 1988; Maestripieri et al., 1999; Rajecki et al., 1978; Salzen, 1970; Sullivan, Landers, et al., 2000). An evolutionary explanation that we have provided for this paradoxical attachment is that it is better for an altricial infant to have a bad caretaker than no caretaker, as an altricial infant is dependent upon access to the mother's milk, warmth, and protection (Hofer & Sullivan, 2001).

However, it is important to discuss the data that demonstrate that infants can learn aversions under some circumstances. Infant rats are able to learn to avoid odors if these are paired with malaise, such as that produced by a LiCl injection or 1.0-mA shock (Abate, Spear, & Molina, 2001; Alleva & Calamandrei, 1986; Campbell, 1984; Coopersmith, Lee, & Leon, 1986; Gruest, Richer, & Hars, 2004; Haroutunian & Campbell, 1979; J. W. Hennessy, Smotherman, & Levine, 1976; Hoffmann, Hunt, & Spear, 1990; Hoffmann, Molina, Kucharski, & Spear, 1987; Hunt, Molina, Rajachandran, Spear, & Spear, 1993; Hunt, Spear, & Spear, 1991; Miller, Molina, & Spear, 1990; Molina, Hoffmann, & Spear, 1986; Richardson & McNally, 2003; Rudy & Cheatle, 1983; Shionoya et al., 2006; Smotherman, 1982; Smotherman, Hennessy, & Levine, 1976; Smotherman & Robinson, 1985, 1990; Spear, 1978; Spear & Rudy, 1991; Stickrod, Kimble, & Smotherman, 1982). Interestingly, while in adult and preweaning rats, the amygdala responds to odor-malaise conditioning (Bermudez-Rattoni, Grijalva, Kiefer, & Garcia, 1986; Gale et al., 2004; LeDoux, 2000; Touzani & Sclafani, 2005), in infants, odor-malaise uses a nonamygdala neural circuit for odor aversion learning that includes the olfactory bulb (Raineki et al., 2009; Shionoya et al., 2006). Another remarkable constraint exists on aversion learning during infancy: If neonatal rats are nursing during odor-LiCl conditioning, this prevents a learned odor aversion and instead produces a learned odor preference (Gubernick & Alberts, 1984;

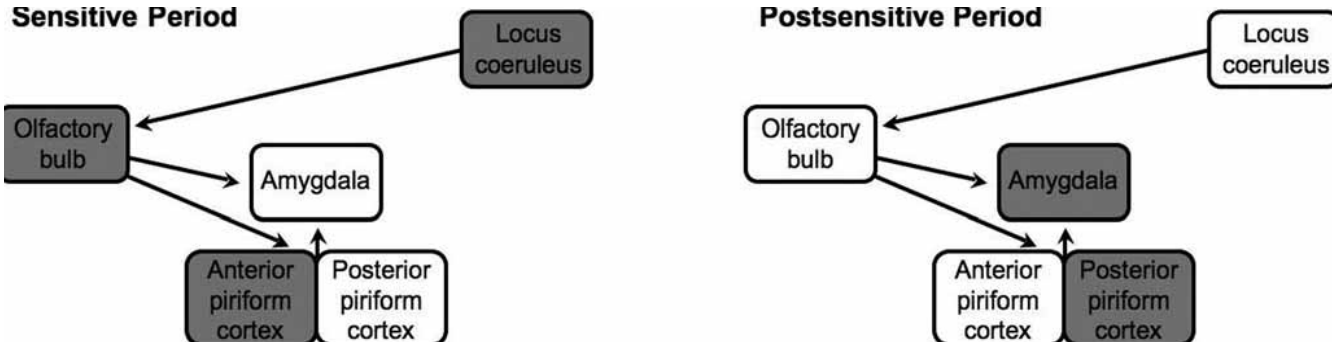


FIGURE 60.2 This graph illustrates the neural basis of attachment learning with odor–0.5-mA shock conditioning during the early life sensitive period for attachment and its maturational changes to the fear neural circuit in older pups. The olfactory bulb, anterior piriform cortex, and LC constitute the attachment neural circuit, while the fear-conditioning neural circuit activates the posterior piriform cortex and the amygdala during fear conditioning in older postsensitive period pups.

Martin & Alberts, 1979; Melcer, Alberts, & Gubernick, 1985; Shionoya et al., 2006).

Together, data indicate that aversions are not readily learned by infants, and we attribute this to unique neural circuitry optimized to facilitate attachment to the caregiver, regardless of the quality of care provided. In figure 60.2, we provide a model of our current understanding of this early social attachment circuit and how this circuitry changes to transition the developing animal from attachment learning to learning that can accommodate both learned preferences and avoidances.

Role of corticosterone in early life

As was discussed in the previous section, the developmental emergence of fear learning parallels amygdala plasticity and maturation (Berdel & Morys, 2000b; Berdel et al., 1997; Bouwmeester et al., 2002; Cunningham et al., 2002; Morys et al., 1999; Nair & Gonzalez-Lima, 1999; Schwob, Haberly, & Price, 1984; Schwob & Price, 1984b; Thompson et al., 2008; Wilson, Best, & Sullivan, 2004). Pharmacological manipulations of corticosterone (CORT) levels in the infant have allowed us to further define the early social circuit, and furthermore have provided us a platform to assess how changes in the early environment can affect the developing brain and subsequent transition to adultlike behavior.

In infant rats, CORT levels are relatively low (Henning, 1978; Walker, Sapolsky, Meaney, Vale, & Rivier, 1986), and the ability of most stressful stimuli, that is, restraint or shock (Grino, Paulmyer-Lacroix, Faudon, Renard, & Anglade, 1994; Levine, 1962, 2001; Rosenfeld, Suchecki, & Levine, 1992), to evoke CORT secretion is greatly reduced in comparison to that in older animals (Butte, Kakihana, Farnham, & Noble, 1973; Cate & Yasumura, 1975; Guillet & Michaelson, 1978; Guillet, Saffran, & Michaelson, 1980; Levine, 1967). This period of reduced hypothalamic-pituitary-adrenal (HPA) axis responsiveness during neonatal development has been termed the *stress hyporesponsive period* (SHRP). Sensory stimulation provided by the mother during nursing and grooming seems to control the pups' low CORT levels (Levine, 1962; Van Oers, Kloet, Whelan, & Levine, 1998). In fact, prolonged maternal separation (~24 hours), which deprives pups of maternal sensory stimulation, increases pups' CORT levels (Levine, 2001), while the replacement of maternal sensory stimulation or maternal presence is able to reinstate the low level of CORT (Stanton & Levine, 1990; Stanton, Wallstrom, & Levine, 1987; Suchecki, Rosenfeld, & Levine, 1993). This reduced stress reactivity experienced by neonates is hypothesized to protect the developing organism from the negative influences of stress hormones (Sapolsky & Meaney, 1986). Indeed, high doses of CORT administrated to the neonatal rat causes decreased mitosis, myelination, and altered granule cell genesis (Bohn, 1980). Furthermore, animals treated during infancy with CORT show reduced DNA content and brain size as well as impaired adult behavior (Bohn, 1984) and neuroendocrine function (Erkine, Geller, & Yuwiler, 1979). But it is important to note that moderate exposure to CORT during the developmental stage may be beneficial. For example, juvenile rats who were exposed to CORT via the dam's milk show superior performance on the Morris water maze task, a test of spatial memory (McCornick et al., 2001).

In adolescents and adults, while stress is generally considered to be detrimental to social behavior, in moderation it has an adaptive role and facilitates social interactions, learning, and the expression of learned social behavior (DeVries, 2002; McEwen, 2002). Indeed, social stimuli directly influence the CORT response. Specifically, maternal presence in adolescent guinea pigs, peers in nonhuman primates, and mate presence in voles reduce CORT (Carter & Keverne, 2002; DeVries, Glasper, & Detillion, 2003; M. B. Hennessy, Maken, & Graves, 2002; M. B. Hennessy, Nigh, Sims, & Long, 1995), while social affiliation in humans blocks stress-induced CORT release (Kirschbaum, Klauer, Filipp, & Hellhammer, 1995). Higher stress levels can produce a defensive/offensive system under perceived danger that is

controlled, at least in part, by the amygdala (Korte, 2001). Thus an adult's ability to balance the stress response determines whether social interactions occur or are inhibited by a fear/anxiety response.

Like the role of NE, the role of CORT in mediating learned behavior changes with maturation. While CORT is considered to play a modulatory role in adult fear conditioning (Corodimas, LeDoux, Gold, & Schulkin, 1994; Hui et al., 2004; Pugh, Tremblay, Fleshner, & Rudy, 1997; Roozendaal, Carmi, & McGaugh, 1996; Roozendaal, Quirarte, & McGaugh, 2002; Thompson, Erickson, Schulkin, & Rosen, 2004), CORT is able to switch whether infants learn an aversion or a preference. Specifically, increasing CORT by systemic injections or by intra-amygdala infusions during 0.5-mA odor-shock conditioning or presentation of naturally aversive stimuli is sufficient to elicit both a fear response (learned or unlearned fear) and amygdala participation in the infant (Moriceau et al., 2004; Moriceau & Sullivan, 2004, 2006; Takahashi, 1994). Maternal presence in older animals will lower CORT levels following stressful stimuli such as shock (Stanton et al., 1987; Suchecki et al., 1993), block fear learning, reinstate the attachment learning (preference), and prevent the participation of the amygdala in learning (Moriceau & Sullivan, 2006). After PN15, only fear will be learned during an odor-shock conditioning (Upton et al., in prep). Furthermore, we have verified the causal relationship between maternal presence and suppression of a shock-induced CORT release in pups' odor aversion learning by systemic and intra-amygdala CORT infusions, which then permit pups to learn odor aversions even in the presence of the mother.

To summarize, data indicate that during the attachment period, the mother maintains low infant CORT levels and attenuates amygdala activation, preventing infants from responding to fear/aversive stimuli. Furthermore, through manipulation of CORT levels, we have highlighted a transition period of co-occurrence between the infant attachment learning system and the amygdala-dependent fear learning system (figure 60.3). This suggests that environmentally induced alterations of CORT levels and amygdala activity have the potential to disrupt the learning transition and underlying neural circuitry.

Consequences of early-life alterations in CORT and amygdala activity

The importance of the early environment in the regulation of behavior throughout the life span has been long recognized in both clinical and experimental studies. Indeed, it suffices to say that adult behavior is dependent on the caregiver and the quality of the caregiving environment. In particular, early-life experiences in the context of early social attachment have the most profound impact on adolescent and adult emotion and cognition in rodents, nonhuman primates, and humans (Bell & Denenberg, 1962; Denenberg, 1963; Harlow & Harlow, 1965; Levine, 1962; Rosenzweig et al., 1969; Schore, 2001). For example, the learned attachment odor in rodents is retained and preferred well into adulthood (Coopersmith & Leon, 1986; Fillion & Blass, 1986; Moore et al., 1996; Sevelinges et al., 2007; Shah, Oxley, Lovic, & Fleming, 2002; Woo & Leon, 1988), although the role of the odor in modifying behavior changes from that used during infancy (attachment to the mother) to that used in adulthood (reproduction). Specifically, following odor-stroke attachment learning in infancy, adult male rats exhibit enhanced sexual performance when exposed to the same odors that they experienced in infancy (Fillion & Blass, 1986; Moore et al., 1996).

These results are consistent with observations in other species on the influence of early experiences on adult mate preferences (Slagsvold, Hansen, Johannessen, & Lifjeld, 2002). Infant-learned attachment odors also continue to elicit both enhanced neural responses of the olfactory bulb and attenuated amygdala activation in the adult (Sevelinges et al., 2007). In particular, an odor that is paired with pain to produce the learned attachment odor attenuates adult fear conditioning as well as attenuating amygdala neural activity supporting the learning (Sevelinges et al., 2007).

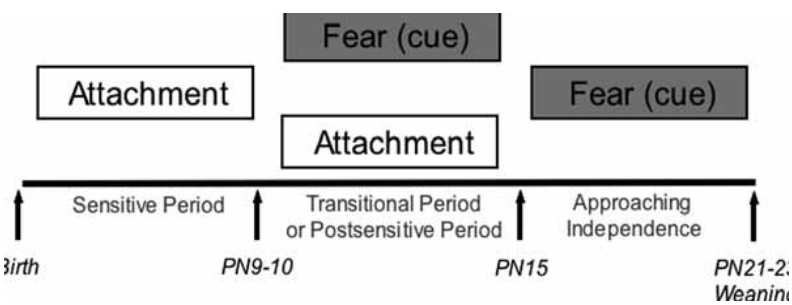


FIGURE 60.3 This schematic represents pups' developmental learning transitions with odor–0.5-mA shock conditioning. Our previous work suggests that PN10 is a transitional age for the onset of amygdala-dependent fear conditioning, although this can be advanced or retarded by increasing or decreasing CORT either pharmacologically or naturally (maternal presence lowers CORT).

Though it is becoming increasingly clear that disruptions to infant attachment have profound maladaptive effects on adult behavior, the question of *how* this occurs remains largely unclear. Different models studying the involvement of early-life environment and its enduring effect have been developed over the years (maternal separation/deprivation, rearing environment alteration and CORT manipulation, neonatal handling), and data from these models in addition to data from our rat model of attachment are providing a clearer understanding of the common loci between infant attachment learning and the damaging effects of early stress on adult behavior.

Maternal Separation/Deprivation The maternal deprivation or maternal separation paradigm is a model of infant neglect. This paradigm, in rats, consists of removing pups from the nest for an extended period of time (3–24 hours) either once or multiple times during the first and second postnatal weeks. Such separation removes multiple sensory stimuli, including odor, warmth, and physical contact, that regulate various aspects of pups' physiology, including CORT, temperature, and heart rate (Hofer, 1973). The difference in removing more or less of these sensory stimuli that are normally provided by the mother, siblings, and the nest has led to variable results between labs.

Rat pups' initial behavioral responses to maternal separation consist of increased behavioral activity and vocalizations, including ultrasonic vocalization (Hofer, Shair, Masmela, & Brunelli, 2001). But within approximately an hour, this response changes to hypoactivity (Hofer & Shair, 1991). Such behavioral responses can be greatly attenuated if pups are provided with adequate warmth and a source of maternal odor (Hofer & Shair, 1978; Sokoloff & Blumberg, 1997). The long-term effects of prolonged periods of maternal separation appear to produce an animal that is more behaviorally responsive to stressful situations (Andersen, Lyss, Dumount, & Teicher, 1999; Kosten, Miserendino, Bombace, Lee, & Kim, 2005). Furthermore, in rats and nonhuman primates, maternal separation during infancy increases the magnitude of the neuroendocrine responses to stress and thus the susceptibility to stress-related diseases (Caldji, Diorio, & Meaney, 2003; Hall, Wilkinson, Humby, & Robbins, 1999; Higley, Hasert, Suomi, & Linnoila, 1991; Ladd et al., 2000; Liu, Caldji, Sharma, Plotsky, & Meaney, 2000; Meaney, 2001; Plotsky & Meaney, 1993; Suomi, 1997). For example, rats that are separated from their mother show greater ACTH and CORT peak responses and a more prolonged response to stress (Ladd et al., 2000; Liu et al., 2000). A reduction in hippocampal glucocorticoid receptors (GR) and thus the reduced glucocorticoid negative feedback appears to be responsible for this increased stress response (Meaney et al., 1996).

Rearing Environment Alteration Nest building begins in pregnancy and continues throughout lactation, although nest quality can vary considerably between mothers. Disruption of a mother's nest-building ability thus provides another avenue to study the effects of early stress on the brain and behavior. In this model, insufficient bedding material is used to provide a continuous stressor for the mother and pups, which ultimately alters mother-pup social interactions. The mother spends a longer time away from the pups, transports the pups more frequently, and engages in more self-directed behaviors that are nonmaternal (grooming) but still nurses normally (Avishai-Eliner, Gilles, Eghbal-Ahmadi, Bar-El, & Baram, 2001; Gilles, Schultz, & Baram, 1996).

The chronic nest alteration paradigm not only raises CORT levels in the infant, but also changes gene expression at multiple levels of the HPA axis and within the frontal cortex (Avishai-Eliner et al., 2001; Gilles et al., 1996; Hatalski, Guirquis, & Baram, 1998; Lightman & Harbuz, 1993; van Oers, de Kloet, Li, & Levine, 1998). In the behavioral realm, this early-stress model produces significant deficits in adult hippocampal learning and memory (Brunson et al., 2005; Fenoglio, Brunson, & Baram, 2006). Finally, we have data showing that attachment learning is modified. Specifically, fear conditioning in infants (which normally produces a learned preference) that are raised in an altered nest environment produces odor aversion learning, and this atypical infant behavior is correlated with the early emergence of amygdala participation and increased CORT levels (Moriceau et al., in prep).

Neonatal Handling Early handling is defined as the experimenter picking up the pup, removing it from the home cage, and isolating it in a different environment for 3–15 minutes daily between birth and weaning. Infant-handled rats show reduced fear as expressed by the increased exploratory activity, a decrease CORT response following stressors, and a more rapid return of CORT to the baseline in adulthood (Hess, Denenberg, Zarrow, & Pfeifer, 1969; Levine, 1962, 1967; Meaney et al., 1993; Meerlo, Horvath, Nagy, Bohus, & Koolhaas, 1999). This rapid return is attributable to a more sensitive HPA axis feedback. Indeed, infant-handled rats have an increase in the number of GRs in the hippocampus and frontal cortex as early as PN23 (Avishai-Eliner et al., 2001; Levine, 1994; Meaney et al., 1993; Sapolsky, 1994). Also, there is a decrease in GRs in the central nucleus of the amygdala around PN9 (Fenoglio, Brunson, Avishai-Eliner, Chen, & Baram, 2004). Both the reduction of GR expression in the amygdala and the increase in the hippocampus may function to reduce the sensitivity of the HPA axis to stressors. However, a maladaptive consequence of neonatal handling is its impact on adult social behavior, which includes a decrease in adult sexual

behavior, ovulation, and sperm production (Gomes, Frantz, Sanvitto, Anselmo-Franci, & Lucion, 1999; Gomes et al., 2005; Mazaro & Lamano-Carvalho, 2006; Padoin, Cadore, Gomes, Barros, & Lucion, 2001; Raineki et al., 2008).

Summary and implications

The clinical literature has clearly shown that early-life adverse experiences (physical and/or emotional) can compromise adult mental health and social behavior (Gunnar & Quevedo, 2007; Teicher et al., 2003). The infant's primary environment is the caregiver, and while the environment expands as the child becomes more mobile and independent, the child's primary environmental force remains the caregiver. Clinical literature suggests that the infant's relationship with the caregiver is of the utmost importance in shaping the child's behavior (Gunnar & Quevedo, 2007; Schore, 2001; Teicher et al., 2003). For example, a child with a healthy and secure attachment is likely to mature into a mentally healthy adult, while a child in an abusive situation has a greater probability of experiencing adult mental dysfunction and physical health problems. Indeed, the clinical effects of an abusive relationship inside and outside of the attachment dyad have different clinical outcomes, with greater vulnerability to later mental health problems when the abuse occurs within the attachment system (Zeanah, Keyes, & Settles, 2003). The neurobiological effects of abuse within versus outside the attachment system remain elusive, especially with respect to the specific physical mechanism that causes such differential effects. Most of the clinical work suggests that both mental and physical health is compromised and expressed during childhood and that this continues through adolescence into adulthood (Bremner, 2003; Nemeroff, 2004).

The importance of these clinical studies has recently been highlighted with brain imaging research showing that these early adverse events are correlated with aberrant adult brain functioning, most notably in the limbic system, frontal cortex, and cerebellum (Bremner, 2003; Kaufman, Plotsky, Nemeroff, & Charney, 2000; Nemeroff, 2004; Teicher et al., 2003). Presumably, these changes arise through maltreatment-induced compromises in the trajectory of brain development (Stien & Kendall, 2004). However, owing to ethical and practical issues, functional imaging of the immature human brain is not feasible under most circumstances. These procedures generally require the child to remain motionless, and therefore anesthesia is required. Thus we are left guessing when a particular brain area function emerges on the basis of anatomical, neurotransmitter, and synaptic development. This problem is potentiated by difficulty in assessing connectivity within and between brain areas as well as deciding whether a child's brain area has the same or different function as that area in the adult.

Our rodent animal model of attachment enables us to assess some factors potentially associated with the clinical outcome. This model of attachment accommodates both abusive and pleasant attachment, yields an experimental paradigm in which the effects of both endogenous and exogenous pharmacological insults to the developing brain and behavior can be assessed, and allows us to identify the basic neural circuitry for early social behavior (attachment learning).

In this review, we have outlined the neural circuitry that underlies the infant rat's attachment to the mother, highlighting its predisposition to support proximity-seeking behaviors. We suggest that the infant rat's attachment circuit is due not simply to the absence or immaturity of brain structures but rather to the brain having unique characteristics (LC hyperfunctioning and amygdala hypofunctioning) that enable the infant to survive in the environment unique to infancy. More important, we have discussed how temporal characteristics of attachment can be manipulated by both environmental and physiological factors and how these factors may render the animal vulnerable to maladaptive brain development.

While human children show behavior within the attachment system (proximity seeking, tolerance of pain) remarkably similar to that of other species (rat, dog, and nonhuman primate), it is unclear whether this attachment circuitry exists in human infants. However, it does suggest that the human infant's brain is *likely* organized to ensure rapid, robust attachment to their caregiver. This further suggests that environmental and physiological factors may likewise alter attachment and adult emotional and cognitive well-being through disruption of the brain areas involved in the early attachment process.

ACKNOWLEDGMENT This work was supported by grants NICHD-HD33402, NIMH H80603, and NSF-IBN0117234.

REFERENCES

Abate, P., Spear, N. E., & Molina, J. C. (2001). Fetal and infantile alcohol-mediated associative learning in the rat. *Alcohol. Clin. Exp. Res.*, *25*(7), 989–998.

Alberts, J. R., & May, B. (1984). Nonnutritive, thermotactile induction of filial huddling in rat pups. *Dev. Psychobiol.*, *17*(2), 161–181.

Alleva, E., & Calamandrei, G. (1986). Odor-aversion learning and retention span in neonatal mouse pups. *Behav. Neural Biol.*, *46*(3), 348–357.

Andersen, S. L., Lyss, P. J., Dumount, N. L., & Teicher, M. H. (1999). Enduring neurochemical effects of early maternal separation on limbic structures. *Ann. NY Acad. Sci.*, *877*, 756–759.

Armstrong, C. M., DeVito, L. M., & Cleland, T. A. (2006). One-trial associative odor learning in neonatal mice. *Chem. Senses*, *31*(4), 343–349.

Avishai-Eliner, S., Gilles, E., Eghbal-Ahmadi, M., Bar-El, Y., & Baram, T. (2001). Altered regulation of gene and protein

expression of hypothalamic-pituitary-adrenal axis components in an immature rat model of chronic stress. *J. Neuroendocrinol.*, *13*(9), 799–807.

Barr, G. A. (1995). Ontogeny of nociception and antinociception. *NIDA Res. Monogr.*, *158*, 172–201.

Bayer, S. A. (1980). Quantitative 3H-thymidine radiographic analyses of neurogenesis in the rat amygdala. *J. Comp. Neurol.*, *194*(4), 845–875.

Bekinschtein, P., Cammarota, M., Izquierdo, I., & Medina, J. H. (2008). BDNF and memory formation and storage. *Neuroscientist*, *14*(2), 147–156.

Bell, R. W., & Denenberg, V. H. (1962). The interrelationships of shock and critical periods in infancy as they affect adult learning and activity. *Anim. Behav.*, *11*, 21–27.

Berdel, B., & Morys, J. (2000a). Expression of calbindin-D28k and parvalbumin during development of rat's basolateral amygdaloid complex. *Int. J. Dev. Neurosci.*, *18*(6), 501–513.

Berdel, B., & Morys, J. (2000b). Expression of calbindin-D28k and parvalbumin during development of rat's basolateral amygdaloid complex. *Int. J. Dev. Neurosci.*, *18*(6), 501–513.

Berdel, B., Morys, J., & Maciejewska, B. (1997). Neuronal changes in the basolateral complex during development of the amygdala of the rat. *Int. J. Dev. Neurosci.*, *15*(6), 755–765.

Bermudez-Rattoni, F., Grijalva, C. V., Kiefer, S. W., & Garcia, J. (1986). Flavor-illness aversions: The role of the amygdala in the acquisition of taste-potentiated odor aversions. *Physiol. Behav.*, *38*(4), 503–508.

Blair, H. T., Schafe, G. E., Bauer, E. P., Rodrigues, S. M., & LeDoux, J. E. (2001). Synaptic plasticity in the lateral amygdala: A cellular hypothesis of fear conditioning. *Learn. Memory*, *8*(5), 229–242.

Blass, E. M., & Teicher, M. H. (1980). Suckling. *Science*, *210*(4465), 15–22.

Blozovski, D., & Cudennec, A. (1980). Passive avoidence learning in the young rat. *Dev. Psychobiol.*, *13*(5), 513–518.

Bohn, M. C. (1980). Granule cell genesis in the hippocampus of rats treated neonatally with hydrocortisone. *Neuroscience*, *5*(11), 2003–2012.

Bohn, M. C. (1984). Role of glucocorticoids in expression and development of phenylethanolamine *N*-methyltransferase (PNMT) in cells derived from the neural crest: A review. *Psychoneuroendocrinology*, *8*(4), 381–390.

Bouwmeester, H., Smits, K., & Van Ree, J. M. (2002). Neonatal development of projections to the basolateral amygdala from prefrontal and thalamic structures in rat. *J. Comp. Neurol.*, *450*(3), 241–255.

Bowlby, J. (1969). *Attachment and loss* (Vol. 1). New York: Basic Books.

Brake, S. C. (1981). Suckling infant rats learn a preference for a novel olfactory stimulus paired with milk delivery. *Science*, *211*(4481), 506–508.

Bremner, J. D. (2003). Long-term effects of childhood abuse on brain and neurobiology. *Child Adolesc. Psychiatr. Clin. N. Am.*, *12*(2), 271–292.

Brunjes, P. C., & Alberts, J. R. (1979). Olfactory stimulation induces filial preferences for huddling in rat pups. *J. Comp. Physiol. Psychol.*, *93*(3), 548–555.

Brunson, K. L., Kramar, E., Lin, B., Chen, Y., Colgin, L. L., Yanagihara, T. K., et al. (2005). Mechanisms of late-onset cognitive decline after early-life stress. *J. Neurosci.*, *25*(41), 9328–9338.

Butte, J. C., Kakihana, R., Farnham, M. L., & Noble, E. P. (1973). The relationship between brain and plasma corticosterone stress response in developing rats. *Endocrinology*, *92*(6), 1775–1779.

Caldji, C., Diorio, J., & Meaney, M. J. (2003). Variations in maternal care alter GABA(A) receptor subunit expression in brain regions associated with fear. *Neuropsychopharmacology*, *28*(11), 1950–1959.

Camp, L. L., & Rudy, J. W. (1988). Changes in the categorization of appetitive and aversive events during postnatal development of the rat. *Dev. Psychobiol.*, *21*(1), 25–42.

Campbell, B. A. (1984). Reflections on the ontogeny of learning and memory. In R. Kail & N. E. Spear (Eds.), *Comparative perspectives on the development of memory* (pp. 23–35). Hillsdale, NJ: Lawrence Erlbaum.

Carew, T. J. (1996). Molecular enhancement of memory formation. *Neuron*, *16*, 5–8.

Carew, T. J., & Sutton, M. A. (2001). Molecular stepping stones in memory consolidation. *Nat. Neurosci.*, *4*, 769–771.

Carter, C. S., & Keverne, E. B. (2002). The neurobiology of social affiliation and pair bonding. In D. W. Pfaff (Ed.), *Hormones, brain, and behavior* (pp. 299–337). San Diego: Academic Press.

Cate, T. E., & Yasumura, S. (1975). Effects of ACTH and histamine stress on serum corticosterone and adrenal cyclic AMP in immature rats. *Endocrinology*, *96*, 1044–1047.

Collier, A., Mast, J., Meyer, D., & Jacobs, C. (1979). Approach-avoidance conflict in preweanling rats: A developmental study. *Anim. Learn. Behav.*, *7*, 514–520.

Collier, A. C., & Bolles, R. C. (1980). The ontogenesis of defensive reactions to shock in preweanling rats. *Dev. Psychobiol.*, *13*(2), 141–150.

Coopersmith, R., Lee, S., & Leon, M. (1986). Olfactory bulb responses after odor aversion learning by young rats. *Brain Res.*, *389*(1–2), 271–277.

Coopersmith, R., & Leon, M. (1986). Enhanced neural response by adult rats to odors experienced early in life. *Brain Res.*, *371*(2), 400–403.

Coppola, D. M., Coltrane, J. A., & Arsov, I. (1994). Retronasal or internasal olfaction can mediate odor-guided behaviors in newborn mice. *Physiol. Behav.*, *56*(4), 729–736.

Corodimas, K. P., LeDoux, J. E., Gold, P. W., & Schulkin, J. (1994). Corticosterone potentiation of conditioned fear in rats. *Ann. NY Acad. Sci.*, *746*, 392–393.

Cunningham, M. G., Bhattacharyya, S., & Benes, F. M. (2002). Amygdalo-cortical sprouting continues into early adulthood: Implications for the development of normal and abnormal function during adolescence. *J. Comp. Neurol.*, *453*(2), 116–130.

Davis, M. (1997). Neurobiology of fear responses: The role of the amygdala. *J. Neuropsychiatry Clin. Neurosci.*, *9*, 382–402.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.

Denenberg, V. H. (1963). Early experience and emotional development. *Sci. Am.*, *208*, 138–146.

DeVries, A. C. (2002). Interaction among social environment, the hypothalamic-pituitary-adrenal axis, and behavior. *Horm. Behav.*, *41*(4), 405–413.

DeVries, A. C., Glasper, E. R., & Detillion, C. E. (2003). Social modulation of stress responses. *Physiol. Behav.*, *79*(3), 399–407.

Distel, H., & Hudson, R. (1985). The contribution of the olfactory and tactile modalities to the nipple-search behaviour of newborn rabbits. *J. Comp. Physiol. A, 157*(5), 599–605.

Emerich, D. F., Scalzo, F. M., Enters, E. K., Spear, N. E., & Spear, L. P. (1985). Effects of 6-hydroxydopamine-induced

catecholamine depletion on shock-precipitated wall climbing of infant rat pups. *Dev. Psychobiol.*, *18*(3), 215–227.

Erkine, M. S., Geller, E., & Yuwiler, A. (1979). Effects of neonatal hydrocortisone treatment on pituitary and adrenocortical responses to stress in young rats. *Neuroendocrinology*, *29*, 191–199.

Fanselow, M. S., & Gale, G. D. (2003). The amygdala, fear, and memory. *Ann. NY Acad. Sci.*, *985*, 125–134.

Fanselow, M. S., & LeDoux, J. E. (1999). Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron*, *23*(2), 229–232.

Fenoglio, K. A., Brunson, K. L., Avishai-Eliner, S., Chen, Y., & Baram, T. Z. (2004). Region-specific onset of handling-induced changes in corticotropin-releasing factor and glucocorticoid receptor expression. *Endocrinology*, *145*(6), 2702–2706.

Fenoglio, K. A., Brunson, K. L., & Baram, T. Z. (2006). Hippocampal neuroplasticity induced by early-life stress: Functional and molecular aspects. *Front. Neuroendocrinol.*, *27*(2), 180–192.

Ferry, B., & McGaugh, J. L. (2000). Role of amygdala norepinephrine in mediating stress hormone regulation in memory storage. *Acta Pharmacol. Sin.*, *21*(6), 481–493.

Fifer, W., & Moon, C. (1995). The effects of fetal experience with sound. In J. Lecanuet, W. Fifer, N. Krasnegor, & W. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 351–368). Hillsdale, NJ: Lawrence Erlbaum.

Fillion, T., & Blass, E. (1986). Infantile experience with suckling odors determines adult sexual behavior in male rats. *Science*, *231*(4739), 729–731.

Fitzgerald, M. (2005). The development of nociceptive circuits. *Nat. Rev. Neurosci.*, *6*, 507–520.

Fleming, A. S., O'Day, D. H., & Kraemer, G. W. (1999). Neurobiology of mother-infant interactions: Experience and central nervous system plasticity across development and generations. *Neurosci. Biobehav. Rev.*, *23*(5), 673–685.

Foote, S. L., Aston-Jones, G., & Bloom, F. E. (1980). Impulse activity of locus coeruleus neurons in awake rats and monkeys is a function of sensory stimulation and arousal. *Proc. Natl. Acad. Sci. USA*, *77*(5), 3033–3037.

Gale, G. D., Anagnostaras, S. G., Godsil, B. P., Mitchell, S., Nozawa, T., Sage, J. R., et al. (2004). Role of the basolateral amygdala in the storage of fear memories across the adult lifetime of rats. *J. Neurosci.*, *24*(15), 3810–3815.

Galef, B. G., Jr., & Kaner, H. C. (1980). Establishment and maintenance of preference for natural and artificial olfactory stimuli in juvenile rats. *J. Comp. Physiol. Psychol.*, *94*(4), 588–595.

Galef, B. G., Jr., & Sherry, D. F. (1973). Mother's milk: A medium for transmission of cues reflecting the flavor of mother's diet. *J. Comp. Physiol. Psychol.*, *83*(3), 374–378.

Gilles, E. E., Schultz, L., & Baram, T. Z. (1996). Abnormal corticosterone regulation in an immature rat model of continuous chronic stress. *Pediatr. Neurol.*, *15*(2), 114–119.

Goldman, P. S., & Tobach, E. (1967). Behaviour modification in infant rats. *Anim. Behavi.*, *15*(4), 559–562.

Gomes, C. M., Frantz, P. J., Sanvitto, G. L., Anselmo-Franci, J. A., & Lucion, A. B. (1999). Neonatal handling induces anovulatory estrous cycles in rats. *Braz. J. Med. Biol. Res.*, *32*(10), 1239–1242.

Gomes, C. M., Raineki, C., Ramos de Paula, P., Severino, G. S., Helena., C. V. V., Anselmo-Franci, J. A., et al. (2005). Neonatal handling and reproductive function in female rats. *Endocrinology*, *184*, 435–445.

Grino, M., Paulmyer-Lacroix, O., Faudon, M., Renard, M., & Anglade, G. (1994). Blockade of alpha 2-adrenoceptors stimulates basal and stress-induced adrenocorticotropin secretion in the developing rat through a central mechanism independent from corticotropin-releasing factor and arginine vasopressin. *Endocrinology*, *135*(6), 2549–2557.

Gruest, N., Richer, P., & Hars, B. (2004). Emergence of long-term memory for conditioned aversion in the rat fetus. *Dev. Psychobiol.*, *44*(3), 189–198.

Gubernick, D. J., & Alberts, J. R. (1984). A specialization of taste aversion learning during suckling and its weaning-associated transformation. *Dev. Psychobiol.*, *17*(6), 613–628.

Guillet, R., & Michaelson, S. M. (1978). Corticotropin responsiveness in the neonatal rat. *Neuroendocrinology*, *27*(3–4), 119–125.

Guillet, R., Saffran, M., & Michaelson, S. M. (1980). Pituitary-adrenal response in neonatal rats. *Endocrinology*, *106*(3), 991–994.

Gunnar, M., & Quevedo, K. (2007). The neurobiology of stress and development. *Annu. Rev. Psychol.*, *58*, 145–173.

Haberly, L. B. (2001). Parallel-distributed processing in olfactory cortex: New insights from morphological and physiological analysis of neuronal circuitry. *Chem. Senses*, *26*(5), 551–576.

Hall, F. S., Wilkinson, L. S., Humby, T., & Robbins, T. W. (1999). Maternal deprivation of neonatal rats produces enduring changes in dopamine function. *Synapse*, *32*(1), 37–43.

Harley, C. W., & Sara, S. J. (1992). Locus coeruleus bursts induced by glutamate trigger delayed perforant path spike amplitude potentiation in dentate gyrus. *Exp. Brain Res.*, *89*(3), 581–587.

Harlow, H., & Harlow, M. (1965). The affectional system. In A. Schrier, H. Harlow & F. Stollnitz (Eds.), *Behavior of nonhuman primates* (Vol. 2). New York: Academic Press.

Haroutunian, V., & Campbell, B. A. (1979). Emergence of interoceptive and exteroceptive control of behavior in rats. *Science*, *205*(4409), 927–929.

Hatalski, C. G., Guirquis, C., & Baram, T. Z. (1998). Corticotropin releasing factor mRNA expression in the hypothalamic paraventricular nucleus and the central nucleus of the amygdala is modulated by repeated acute stress in the immature rat. *J. Neuroendocrinol.*, *10*(9), 663–669.

Helfer, M. E., Kempe, R. S., & Krugman, R. D. (1997). *The battered child.* Chicago: University of Chicago Press.

Hennessy, J. W., Smotherman, W. P., & Levine, S. (1976). Conditioned taste aversion and the pituitary-adrenal system. *Behav. Biol.*, *16*(4), 413–424.

Hennessy, M. B., Li, J., & Levine, S. (1980). Infant responsiveness to maternal cues in mice of 2 inbred lines. *Dev. Psychobiol.*, *13*(1), 77–84.

Hennessy, M. B., Maken, D. S., & Graves, F. C. (2002). Presence of mother and unfamiliar female alters levels of testosterone, progesterone, cortisol, adrenocorticotropin, and behavior in maturing Guinea pigs. *Horm. Behav.*, *42*(1), 42–52.

Hennessy, M. B., Nigh, C. K., Sims, M. L., & Long, S. J. (1995). Plasma cortisol and vocalization responses of postweaning age guinea pigs to maternal and sibling separation: Evidence for filial attachment after weaning. *Dev. Psychobiol.*, *28*(2), 103–115.

Henning, S. J. (1978). Plasma concentrations of total and free corticosterone during development in the rat. *Am. J. Physiol.*, *235*(5), E451–E456.

Herzog, C., & Otto, T. (1997). Odor-guided fear conditioning in rats: 2. Lesions of the anterior perirhinal cortex disrupt fear conditioned to the explicit conditioned stimulus but not to the training context. *Behav. Neurosci.*, *111*(6), 1265–1272.

Hess, E. (1962). Ethology: An approach to the complete analysis of behavior. In R. Brown, E. Galanter, E. Hess, & G. Mendler

(Eds.), *New directions in psychology* (pp. 159–199). New York: Holt, Rinehart & Winston.

Hess, J. L., Denenberg, V. H., Zarrow, M. X., & Pfeifer, W. D. (1969). Modification of the corticosterone response curve as a function of handling in infancy. *Physiol. Behav.*, *4*(1), 109–111.

Higley, J. D., Hasert, M. F., Suomi, S. J., & Linnoila, M. (1991). Nonhuman primate model of alcohol abuse: Effects of early experience, personality, and stress on alcohol consumption. *Proc. Natl. Acad. Sci. USA*, *88*(16), 7261–7265.

Hofer, M. A. (1973). Studies on how early maternal separation produces behavioral change in young rats. *Psychosom. Med.*, *37*(2), 180–184.

Hofer, M. A., & Shair, H. N. (1978). Ultrasonic vocalization during social interaction and isolation in 2-week-old rats. *Dev. Psychobiol.*, *11*(5), 495–504.

Hofer, M. A., & Shair, H. N. (1991). Independence of ultrasonic vocalization and thermogenic responses in infant rats. *Behav. Neurosci.*, *105*(1), 41–48.

Hofer, M. A., Shair, H. N., Masmela, J. R., & Brunelli, S. A. (2001). Developmental effects of selective breeding for an infantile trait: The rat pup ultrasonic isolation call. *Dev. Psychobiol.*, *39*(4), 231–246.

Hofer, M. A., Shair, H., & Singh, P. (1976). Evidence that maternal ventral skin substances promote suckling in infant rats. *Physiol. Behav.*, *17*(1), 131–136.

Hofer, M. A., & Sullivan, R. M. (2001). Toward a neurobiology of attachment. In C. A. Nelson & M. Lucian (Eds.), *Developmental cognitive neuroscience* (pp. 599–616). Cambridge, MA: MIT Press.

Hoffmann, H., Hunt, P., & Spear, N. E. (1990). Ontogenetic differences in the association of gustatory and tactile cues with lithium chloride and footshock. *Behav. Neural Biol.*, *53*(3), 441–450.

Hoffmann, H., Molina, J. C., Kucharski, D., & Spear, N. E. (1987). Further examination of ontogenetic limitations on conditioned taste aversion. *Dev. Psychobiol.*, *20*(4), 455–463.

Hudson, R. (1985). Do newborn rabbits learn the odor stimuli releasing nipple-search behavior? *Dev. Psychobiol.*, *18*(6), 575–585.

Hudson, R., & Distel, H. (1983). Nipple location by newborn rabbits: Behavioral evidence for pheromonal guidance. *Behaviour*, *85*, 260–275.

Hui, G. K., Figueroa, I. R., Poytress, B. S., Roozendaal, B., McGaugh, J. L., & Weinberger, N. M. (2004). Memory enhancement of classical fear conditioning by post-training injections of corticosterone in rats. *Neurobiol. Learn. Mem.*, *81*(1), 67–74.

Hunt, P. S., Molina, J. C., Rajachandran, L., Spear, L. P., & Spear, N. E. (1993). Chronic administration of alcohol in the developing rat: Expression of functional tolerance and alcohol olfactory aversions. *Behav. Neural Biol.*, *59*(2), 87–99.

Hunt, P. S., Spear, L. P., & Spear, N. E. (1991). An ontogenetic comparison of ethanol-mediated taste aversion learning and ethanol-induced hypothermia in preweanling rats. *Behav. Neurosci.*, *105*(6), 971–983.

Johanson, I. B., & Hall, W. G. (1979). Appetitive learning in 1-day-old rat pups. *Science*, *205*(4404), 419–421.

Johanson, I. B., & Teicher, M. H. (1980). Classical conditioning of an odor preference in 3-day-old rats. *Behav. Neural Biol.*, *29*(1), 132–136.

Johnson, B. A., Woo, C. C., Duong, H., Nguyen, V., & Leon M. (1995). A learned odor evokes an enhanced Fos-like glomerular response in the olfactory bulb of young rats. *Brain Res., 699*(2), 192–200.

Kaufman, J., Plotsky, P. M., Nemeroff, C. B., & Charney, D. S. (2000). Effects of early adverse experiences on brain structure and function: Clinical implications. *Biol. Psychiatry*, *48*(8), 778–790.

Kimura, F., & Nakamura, S. (1985). Locus coeruleus neurons in the neonatal rat: Electrical activity and responses to sensory stimulation. *Brain Res.*, *355*(2), 301–305.

Kirschbaum, C., Klauer, T., Filipp, S. H., & Hellhammer, D. H. (1995). Sex-specific effects of social support on cortisol and subjective responses to acute psychological stress. *Psychosom. Med.*, *57*(1), 23–31.

Korte, S. M. (2001). Corticosteroids in relation to fear, anxiety and psychopathology. *Neurosci. Biobehav. Rev.*, *25*(2), 117–142.

Kosten, T. A., Miserendino, M. J. D., Bombace, J. C., Lee, H. J., & Kim, J. J. (2005). Sex-selective effects of neonatal isolation on fear conditioning and foot shock sensitivity. *Behav. Brain Res.*, *157*(2), 235–244.

Ladd, C. O., Huot, R. L., Thrivikraman, K. V., Nemeroff, C. B., Meaney, M. J., & Plotsky, P. M. (2000). Long-term behavioral and neuroendocrine adaptations to adverse early experience. *Prog. Brain Res.*, *122*, 81–103.

Langdon, P. E., Harley, C. W., & McLean, J. H. (1997). Increased beta adrenoceptor activation overcomes conditioned olfactory learning deficits induced by serotonin depletion. *Dev. Brain Res.*, *102*(2), 291–293.

Lecanuet, J. P., & Schaal, B. (1996). Fetal sensory competencies. *Eur. J. Obstet., Gynecol. Reprod. Biol.*, *68*(1–2), 1–23.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.*, *23*, 155–184.

Leon, M. (1975). Dietary control of maternal pheromone in the lactating rat. *Physiol. Behav.*, *14*, 311–319.

Leon, M. (1992). The neurobiology of filial learning. *Annu. Rev. Psychol.*, *43*, 377–398.

Levine, S. (1962). Plasma-free corticosteroid response to electric shock in rats stimulated in infancy. *Science*, *135*, 795–796.

Levine, S. (1967). Maternal and environmental influences on the adrenocortical response to stress in weanling rats. *Science*, *156*(772), 258–260.

Levine, S. (1994). The ontogeny of the hypothalamic-pituitary-adrenal axis. The influence of maternal factors. *Ann. NY Acad. Sci.*, *746*, 275–288.

Levine, S. (2001). Primary social relationships influence the development of the hypothalamic-pituitary-adrenal axis in the rat. *Physiol. Behav.*, *73*(3), 255–260.

Lightman, S. L., & Harbuz, M. S. (1993). Expression of corticotropin-releasing factor mRNA in response to stress. *Ciba Found. Symp.*, *172*, 173–187; discussion 187–198.

Liu, D., Caldji, C., Sharma, S., Plotsky, P., & Meaney, M. (2000). Influence of neonatal rearing conditions on stress-induced adrenocorticotropin responses and norepinepherine release in the hypothalamic paraventricular nucleus. *J. Neuroendocrinol.*, *12*(1), 5–12.

Maestripieri, D., Tomaszycki, M., & Carroll, K. A. (1999). Consistency and change in the behavior of rhesus macaque abusive mothers with successive infants. *Dev. Psychobiol.*, *34*(1), 29–35.

Maren, S. (2003). The amygdala, synaptic plasticity, and fear memory. *Ann. NY Acad. Sci.*, *985*, 106–113.

Marshall, K. C., Christie, M. J., Finlayson, P. G., & Williams, J. T. (1991). Developmental aspects of the locus coeruleusnoradrenaline system. *Prog. Brain Res.*, *88*, 173–185.

Martin, L. T., & Alberts, J. R. (1979). Taste aversions to mother's milk: The age-related role of nursing in acquisition and

expression of a learned association. *J. Comp. Physiol. Psychol.*, *93*(3), 430–445.

Mazaro, R., & Lamano-Carvalho, T. L. (2006). Prolonged deleterious effects of neonatal handling on reproductive parameters of pubertal male rats. *Reprod. Fertil. Dev.*, *18*(4), 497–500.

McCornick, C. M., Rioux, T., Fisher, R., Lang, K., MacLaury, K., & Teillon, S. M. (2001). Effects of neonatal corticosterone treatment on maze performance and HPA axis in juvenile rats. *Physiol. Behav.*, *74*, 371–379.

McEwen, B. S. (2002). Protective and damaging effects of stress mediators: The good and bad sides of the response to stress. *Metabolism*, *51*(6, Suppl. 1), 2–4.

McGaugh, J. L. (2006). Make mild moments memorable: Add a little arousal. *Trends Cogn. Sci.*, *10*(8), 345–347.

McGaugh, J. L., Roozendaal, B., & Cahill, L. (1999). Modulation of memory storage by stress hormones and the amygdaloid complex. In M. Gazzaniga (Ed.), *Cognitive neuroscience* (2nd ed.). Cambridge, MA: MIT Press.

McLean, J. H., Darby-King, A., Sullivan, R. M., & King, S. R. (1993). Serotonergic influence on olfactory learning in the neonate rat. *Behav. Neural Biol.*, *60*(2), 152–162.

McLean, J. H., Harley, C. W., Darby-King, A., & Yuan, Q. (1999). pCREB in the neonate rat olfactory bulb is selectively and transiently increased by odor preference-conditioned training. *Learn. Memory*, *6*(6), 608–618.

McLean, J. H., & Shipley, M. T. (1991). Postnatal development of the noradrenergic projection from locus coeruleus to the olfactory bulb in the rat. *J. Comp. Neurol.*, *304*(3), 467–477.

Meaney, M. J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annu. Rev. Neurosci.*, *24*, 1161–1192.

Meaney, M. J., Bhatnagar, S., Diorio, J., Larocque, S., Francis, D., O'Donnell, D., et al. (1993). Molecular basis for the development of individual differences in the hypothalamic-pituitary-adrenal stress response. *Cell. Mol. Neurobiol.*, *13*(4), 321–347.

Meaney, M. J., Diorio, J., Francis, D., Widdwson, J., LaPlante, P., Caldji, C., et al. (1996). Early environmental regulation of forebrain glucocorticoid receptor gene expression: Implications for adrenocortical responses to stress. *Dev. Neurosci.*, *18*(1–2), 49–72.

Meerlo, P., Horvath, K. M., Nagy, G. M., Bohus, B., & Koolhaas, J. M. (1999). The influence of postnatal handling on adult neuroendocrine and behavioural stress reactivity. *J. Neuroendocrinol.*, *11*(12), 925–933.

Melcer, T., Alberts, J. R., & Gubernick, D. J. (1985). Early weaning does not accelerate the expression of nursing-related taste aversions. *Dev. Psychobiol.*, *18*(5), 375–381.

Mennella, J. A., Johnson, A., & Beauchamp, G. K. (1995). Garlic ingestion by pregnant women alters the odor of amniotic fluid. *Chem. Senses*, *20*(2), 207–209.

Miller, J. S., Jagielo, J. A., & Spear, N. E. (1989). Age-related differences in short-term retention of separable elements of an odor aversion. *J. Exp. Psychol. [Anim. Behav.]*, *15*(3), 194–201.

Miller, J. S., Molina, J. C., & Spear, N. E. (1990). Ontogenetic differences in the expression of odor-aversion learning in 4- and 8-day-old rats. *Dev. Psychobiol.*, *23*(4), 319–330.

Mizukawa, K., Tseng, I. M., & Otsuka, N. (1989). Quantitative electron microscopic analysis of postnatal development of zinc-positive nerve endings in the rat amygdala using Timm's sulphide silver technique. *Brain Res. Dev. Brain Res.*, *50*(2), 197–203.

Moles, A., Kieffer, B. L., & D'Amato, F. R. (2004). Deficit in attachment behavior in mice lacking the mu-opioid receptor gene. *Science*, *304*(5679), 1983–1986.

Molina, J. C., Hoffmann, H., & Spear, N. E. (1986). Conditioning of aversion to alcohol orosensory cues in 5- and 10-day rats: Subsequent reduction in alcohol ingestion. *Dev. Psychobiol.*, *19*(3), 175–183.

Moore, C., Jordan, L., & Wong, L. (1996). Early olfactory experience, novelty, and choice of sexual partner by male rats. *Physiol. Behav.*, *60*(5), 1361–1367.

Moriceau, S., Roth, T. L., Okotoghaide, T., & Sullivan, R. M. (2004). Corticosterone controls the developmental emergence of fear and amygdala function to predator odors in infant rat pups. *Int. J. Dev. Neurosci.*, *22*(5–6), 415–422.

Moriceau, S., & Sullivan, R. M. (2004). Unique neural circuitry for neonatal olfactory learning. *J. Neurosci.*, *24*(5), 1182–1189.

Moriceau, S., & Sullivan, R. M. (2006). Maternal presence serves as a switch between learning fear and attraction in infancy. *Nat. Neurosci.*, *9*(8), 1004–1006.

Moriceau, S., Wilson, D. A., Levine, S., & Sullivan, R. M. (2006). Dual circuitry for odor-shock conditioning during infancy: Corticosterone switches between fear and attraction via amygdala. *J. Neurosci.*, *26*(25), 6737–6748.

Morys, J., Berdel, B., Jagalska-Majewska, H., & Luczynska, A. (1999). The basolateral amygdaloid complex: Its development, morphology and functions. *Folia Morphol. (Warsz.)*, *58*(3, Suppl. 2), 29–46.

Myslivecek, J. (1997). Inhibitory learning and memory in newborn rats. *Prog. Neurobiol.*, *53*(4), 399–430.

Nair, H., & Gonzalez-Lima, F. (1999). Extinction of behavior in infant rats: Development of functional coupling between septal, hippocampal, and ventral tegmental regions. *J. Neurosci.*, *19*(19), 8646–8655.

Nakamura, S. T., & Sakaguchi, T. (1990). Development and plasticity of the locus coeruleus: A review of recent physiological and pharmacological experimentation. *Prog. Neurobiol.*, *34*, 505–526.

Nemeroff, C. B. (2004). Neurobiological consequences of childhood trauma. *J. Clin. Psychiatry*, *65*(Suppl. 1), 18–28.

Okutani, F., Kaba, H., Takahashi, S., & Seto, K. (1998). The biphasic effects of locus coeruleus noradrenergic activation on dendrodendritic inhibition in the rat olfactory bulb. *Brain Res.*, *783*(2), 272–279.

Padoin, M. J., Cadore, L. P., Gomes, C. M., Barros, H. M., & Lucion, A. B. (2001). Long-lasting effects of neonatal stimulation on the behavior of rats. *Behav. Neurosci.*, *115*(6), 1332–1340.

Pape, H. C., & Stork, O. (2003). Genes and mechanisms in the amygdala involved in the formation of fear memory. *Ann. NY Acad. Sci.*, *985*, 92–105.

Pare, D., Quirk, G. J., & Ledoux, J. E. (2004). New vistas on amygdala networks in conditioned fear. *J. Neurophysiol.*, *92*(1), 1–9.

Pedersen, P. E., Williams, C. L., & Blass, E. M. (1982). Activation and odor conditioning of suckling behavior in 3-day-old albino rats. *J. Exp. Psychol. [Anim. Behav.]*, *8*(4), 329–341.

Plotsky, P. M., & Meaney, M. J. (1993). Early postnatal experience alters hypothalamic corticotropin-releasing factor (CRF) mRNA, median eminence CRF content and stress-induced release in adult rats. *Mol. Brain Res.*, *18*, 195–200.

Polan, H. J., & Hofer, M. A. (1998). Olfactory preference for mother over home nest shavings by newborn rats. *Dev. Psychobiol.*, *33*(1), 5–20.

Pollak, S. D. (2003). Experience-dependent affective learning and risk for psychopathology in children. *Ann. NY Acad. Sci.*, *1008*, 102–111.

Pugh, C. R., Tremblay, D., Fleshner, M., & Rudy, J. W. (1997). A selective role for corticosterone in contextual-fear conditioning. *Behav. Neurosci.*, *111*(3), 503–511.

Raineki, C., Shionoya, K., Sander, K., & Sullivan, R. M. (2009). Ontogeny of odor-LiCl vs. odor-shock learning: Similar behaviors but divergent ages of functional amygdala emergence. *Learn. Mem.*, *16*(2), 114–121.

Raineki, C., Szawka, R. E., Gomes, C. M., Lucion, M. K., Barp, J., Bello-Klein, A., Franci, C. R., Anselmo-Franci, J. A., & Lucion, A. B. (2008). Effects of neonatal handling on central noradrenergic and nitric oxidergic systems and reproductive parameters in female rats. *Neuroendocrinology*, *87*(3), 151–159.

Rajecki, D., Lamb, M., & Obmascher, P. (1978). Towards a general theory of infantile attachment: A comparative review of aspects of the social bond. *Behav. Brain Sci.*, *3*, 417–464.

Rangel, S., & Leon, M. (1995). Early odor preference training increases olfactory bulb norepinephrine. *Brain Res. Dev. Brain Res.*, *85*(2), 187–191.

Rankin, C. H. (2002). Neuroscience: A bite to remember. *Science*, *296*(5573), 1624–1625.

Richardson, R., & McNally, G. P. (2003). Effects of an odor paired with illness on startle, freezing, and analgesia in rats. *Physiol. Behav.*, *78*(2), 213–219.

Risser, J. M., & Slotnick, B. M. (1987). Nipple attachment and survival in neonatal olfactory bulbectomized rats. *Physiol. Behav.*, *40*(4), 545–549.

Roozendaal, B., Carmi, O., & McGaugh, J. L. (1996). Adrenocortical suppression blocks the memory-enhancing effects of amphetamine and epinephrine. *Proc. Natl. Acad. Sci. USA*, *93*(4), 1429–1433.

Roozendaal, B., Quirarte, G. L., & McGaugh, J. L. (2002). Glucocorticoids interact with the basolateral amygdala beta-adrenoceptor–cAMP/cAMP/PKA system in influencing memory consolidation. *Eur. J. Neurosci.*, *15*(3), 553–560.

Rosenfeld, P., Suchecki, D., & Levine, S. (1992). Multifactorial regulation of the hypothalamic pituitary-adrenal axis during development. *Neurosci. Biobehav. Rev.*, *16*, 553–568.

Rosenkranz, J. A., & Grace, A. A. (2002). Dopamine-mediated modulation of odour-evoked amygdala potentials during Pavlovian conditioning. *Nature*, *417*(6886), 282–287.

Rosenzweig, M. R., Bennett, E. L., Diamond, M. C., Wu, S. Y., Slagle, R. W., & Saffran, E. (1969). Influences of environmental complexity and visual stimulation on development of occipital cortex in rat. *Brain Res.*, *14*(2), 427–445.

Roth, T., & Sullivan, R. (2005). Memory of early maltreatment: Neonatal behavioral and neural correlates of maternal maltreatment within the context of classical conditioning. *Biol. Psychiatry*, *57*(8), 823–831.

Rudy, J. W., & Cheatle, M. D. (1977). Odor-aversion learning in neonatal rats. *Science*, *198*(4319), 845–846.

Rudy, J. W., & Cheatle, M. D. (1983). Odor-aversion learning by rats following LiCl exposure: Ontogenetic influences. *Dev. Psychobiol.*, *16*(1), 13–22.

Salzen, E. (1970). Imprinting and environmental learning. In E. T. L. Aronson, D. Lehrman, J. Rosenblatt (Ed.), *Development and evolution of behavior* (pp. 158–178). San Francisco: W. H. Freeman.

Sananes, C., & Campbell, B. (1989). Role of the central nucleus of the amygdala in olfactory heart rate conditioning. *Behav. Neurosci.*, *103*(3), 519–525.

Sanchez, M. M., Ladd, C. O., & Plotsky, P. M. (2001). Early adverse experience as a developmental risk factor for later psychopathology: Evidence from rodent and primate models. *Dev. Psychopathol.*, *13*(3), 419–449.

Sapolsky, R. M. (1994). The physiological relevance of glucocorticoid endangerment of the hippocampus. *Ann. NY Acad. Sci.*, *746*, 294–304.

Sapolsky, R. M., & Meaney, M. J. (1986). Maturation of the adrenocortical stress response: Neuroendocrine control mechanisms and the stress hyporesponsive period. *Brain Res.*, *396*(1), 64–76.

Sara, S. J., Dyon-Laurent, C., & Herve, A. (1995). Novelty seeking behavior in the rat is dependent upon the integrity of the noradrenergic system. *Cogn. Brain Res.*, *2*(3), 181–187.

Schaal, B., Marlier, L., & Soussignan, R. (1995). Responsiveness to the odour of amniotic fluid in the human neonate. *Biol. Neonate*, *67*(6), 397–406.

Schettino, L. F., & Otto, T. (2001). Patterns of Fos expression in the amygdala and ventral perirhinal cortex induced by training in an olfactory fear conditioning paradigm. *Behav. Neurosci.*, *115*(6), 1257–1272.

Schore, A. N. (2001). The effects of early relational trauma on right brain development, affect regulation, and infant mental health. *Infant Ment. Health J.*, *22*, 201–269.

Schwob, J. E., Haberly, L. B., & Price, J. L. (1984). The development of physiological responses of the piriform cortex in rats to stimulation of the lateral olfactory tract. *J. Comp. Neurol.*, *223*(2), 223–237.

Schwob, J. E., & Price, J. L. (1984a). The development of axonal connections in the central olfactory system of rats. *J. Comp. Neurol.*, *223*(2), 177–202.

Schwob, J. E., & Price, J. L. (1984b). The development of lamination of afferent fibers to the olfactory cortex in rats, with additional observations in the adult. *J. Comp. Neurol.*, *223*(2), 203–222.

Sevelinges, Y., Gervais, R., Messaoudi, B., Granjon, L., & Mouly, A. M. (2004). Olfactory fear conditioning induces field potential potentiation in rat olfactory cortex and amygdala. *Learn. Memory*, *11*(6), 761–769.

Sevelinges, Y., Moriceau, S., Holman, P., Miner, C., Muzny, K., Gervais, R., et al. (2007). Enduring effects of infant memories: Infant odor-shock conditioning attenuates amygdala activity and adult fear conditioning. *Biol. Psychiatry*, *62*(10), 1070–1079.

Shah, A., Oxley, G., Lovic, V., & Fleming, A. S. (2002). Effects of preweaning exposure to novel maternal odors on maternal responsiveness and selectivity in adulthood. *Dev. Psychobiol.*, *41*(3), 187–196.

Shair, H. N., Masmela, J. R., Brunelli, S. A., & Hofer, M. A. (1997). Potentiation and inhibition of ultrasonic vocalization of rat pups: Regulation by social cues. *Dev. Psychobiol.*, *30*(3), 195–200.

Shionoya, K., Moriceau, S., Lunday, L., Miner, C., Roth, T. L., & Sullivan, R. M. (2006). Development switch in neural circuitry underlying odor-malaise learning. *Learn. Memory*, *13*(6), 801–808.

Shipley, M. T., Halloran, F. J., & de la Torre, J. (1985). Surprisingly rich projection from locus coeruleus to the olfactory bulb in the rat. *Brain Res.*, *329*(1–2), 294–299.

Sigurdsson, T., Doyere, V., Cain, C. K., & LeDoux, J. E. (2007). Long-term potentiation in the amygdala: A cellular mechanism of fear learning and memory. *Neuropharmacology*, *52*(1), 215–227.

Slagsvold, T., Hansen, B. T., Johannessen, L. E., & Lifjeld, J. T. (2002). Mate choice and imprinting in birds studied by cross-fostering in the wild. *Proc. R. Soc. Lond. B Biol. Sci.*, *269*(1499), 1449–1455.

Smotherman, W. P. (1982). Odor aversion learning by the rat fetus. *Physiol. Behav.*, *29*(5), 769–771.

Smotherman, W. P., Hennessy, J. W., & Levine, S. (1976). Plasma corticosterone levels during recovery from LiCl produced taste aversions. *Behav. Biol.*, *16*(4), 401–412.

Smotherman, W. P., & Robinson, S. R. (1985). The rat fetus in its environment: Behavioral adjustments to novel, familiar, aversive, and conditioned stimuli presented in utero. *Behav. Neurosci.*, *99*(3), 521–530.

Smotherman, W. P., & Robinson, S. R. (1990). Rat fetuses respond to chemical stimuli in gas phase. *Physiol. Behav.*, *47*(5), 863–868.

Sokoloff, G., & Blumberg, M. S. (1997). Thermogenic, respiratory, and ultrasonic responses of week-old rats across the transition from moderate to extreme cold exposure. *Dev. Psychobiol.*, *30*(3), 181–194.

Spear, N. (1978). *Processing memories: Forgetting and retention*. Hillsdale, NJ: Lawrence Erlbaum.

Spear, N. E., & Rudy, J. W. (1991). Tests of the ontogeny of learning and memory: Issues, methods, and results. In H. N. Shair, G. A. Barr, et al. (Eds.), *Developmental psychobiology: New methods and changing concepts* (pp. 84–113). New York: Oxford University Press.

Stanley, W. (1962). Differential human handling as reinforcing events and as treatments influencing later social behavior in Basenji puppies. *Psychol. Reports 10*, 775–788.

Stanton, M., & Levine, S. (1990). Inhibition of infant glucocorticoid stress response: Specific role of maternal cues. *Dev. Psychobiol.*, *23*(5), 411–426.

Stanton, M. E., Wallstrom, J., & Levine, S. (1987). Maternal contact inhibits pituitary-adrenal stress responses in preweanling rats. *Dev. Psychobiol.*, *20*(2), 131–145.

Stehouwer, D. J., & Campbell, B. A. (1978). Habituation of the forelimb-withdrawal response in neonatal rats. *J. Exp. Psychol. [Anim. Behav.]*, *4*(2), 104–119.

Stickrod, G., Kimble, D. P., & Smotherman, W. P. (1982). In utero taste/odor aversion conditioning in the rat. *Physiol. Behav.*, *28*(1), 5–7.

Stien, P., & Kendall, J. (2004). *Psychological trauma and the developing brain: Neurologically based interventions for troubled children*. Binghamton, NY: Haworth Press.

Suchecki, D., Rosenfeld, P., & Levine, S. (1993). Maternal regulation of the hypothalamic-pituitary-adrenal axis in the infant rat: The roles of feeding and stroking. *Brain Res.*, *75*(2), 185–192.

Sullivan, R. M., Brake, S. C., Hofer, M. A., & Williams, C. L. (1986). Huddling and independent feeding of neonatal rats can be facilitated by a conditioned change in behavioral state. *Dev. Psychobiol.*, *19*(6), 625–635.

Sullivan, R. M., Hofer, M. A., & Brake, S. C. (1986). Olfactory-guided orientation in neonatal rats is enhanced by a conditioned change in behavioral state. *Dev. Psychobiol.*, *19*(6), 615–623.

Sullivan, R. M., Landers, M., Yeaman, B., & Wilson, D. A. (2000). Good memories of bad events in infancy. *Nature*, *407*(6800), 38–39.

Sullivan, R. M., & Leon, M. (1986). Early olfactory learning induces an enhanced olfactory bulb response in young rats. *Brain Res.*, *392*(1–2), 278–282.

Sullivan, R. M., Stackenwalt, G., Nasr, F., Lemon, C., & Wilson, D. A. (2000). Association of an odor with activation of olfactory bulb noradrenergic beta-receptors or locus coeruleus stimulation is sufficient to produce learned approach responses to that odor in neonatal rats. *Behav. Neurosci.*, *114*(5), 957–962.

Sullivan, R. M., Taborsky-Barba, S., Mendoza, R., Itano, A., Lean, M., Cotman, C., Payne, T., & Lott, I. (1991). Olfactory classical conditioning in neonates. *Pediatrics*, *87*, 511–518.

Sullivan, R. M., & Wilson, D. A. (1991). Neural correlates of conditioned odor avoidance in infant rats. *Behav. Neurosci.*, *105*(2), 307–312.

Sullivan, R. M., Wilson, D. A., Lemon, C., & Gerhardt, G. A. (1994). Bilateral 6-OHDA lesions of the locus coeruleus impair associative olfactory learning in newborn rats. *Brain Res.*, *64*(1–2), 306–309.

Sullivan, R. M., Wilson, D. A., & Leon, M. (1989). Norepinephrine and learning-induced plasticity in infant rat olfactory system. *J. Neurosci.*, *9*(11), 3998–4006.

Sullivan, R. M., Wilson, D. A., Wong, R., Correa, A., & Leon, M. (1990). Modified behavioral and olfactory bulb responses to maternal odors in preweanling rats. *Brain Res. Dev. Brain Res.*, *53*(2), 243–247.

Sullivan, R. M., Zyzak, D. R., Skierkowski, P., & Wilson, D. A. (1992). The role of olfactory bulb norepinephrine in early olfactory learning. *Brain Res. Dev. Brain Res.*, *70*(2), 279–282.

Suomi, S. J. (1997). Early determinants of behaviour: Evidence from primate studies. *Br. Med. Bull.*, *53*(1), 170–184.

Suomi, S. J. (2003). Gene-environment interactions and the neurobiology of social conflict. *Ann. NY Acad. Sci.*, *1008*, 132–139.

Swanson, L. W., & Petrovich, G. D. (1998). What is the amygdala?. *Trends Neurosci.*, *21*(8), 323–331.

Takahashi, L. K. (1994). Organizing action of corticosterone on the development of behavioral inhibition in the preweanling rat. *Brain Res. Dev. Brain Res.*, *81*(1), 121–127.

Tao, X., Finkbeiner, S., Arnold, D. B., Shaywitz, A. J., & Greenberg, M. E. (1998). Ca^{2+} influx regulates BDNF transcription by a CREB family transcription factor-dependent mechanism. *Neuron*, *20*(4), 709–726.

Teicher, M. H., Andersen, S. L., Polcari, A., Anderson, C. M., Navalta, C. P., & Kim, D. M. (2003). The neurobiological consequences of early stress and childhood maltreatment. *Neurosci. Biobehav. Rev.*, *27*(1–2), 33–44.

Teicher, M. H., & Blass, E. M. (1977). First suckling response of the newborn albino rat: The roles of olfaction and amniotic fluid. *Science*, *198*(4317), 635–636.

Teicher, M. H., Flaum, L. E., Williams, M., Eckhert, S. J., & Lumia, A. R. (1978). Survival, growth and suckling behavior of neonatally bulbectomized rats. *Physiol. Behav.*, *21*(4), 553–561.

Terry, L. M., & Johanson, I. B. (1996). Effects of altered olfactory experiences on the development of infant rats' responses to odors. *Dev. Psychobiol.*, *29*(4), 353–377.

Thompson, B. L., Erickson, K., Schulkin, J., & Rosen, J. B. (2004). Corticosterone facilitates retention of contextually conditioned fear and increases CRH mRNA expression in the amygdala. *Behav. Brain Res.*, *149*(2), 209–215.

Thompson, J. V., Sullivan, R. M., & Wilson, D. A. (2008). Developmental emergence of fear learning corresponds with changes in amygdala synaptic plasticity. *Brain Res.*, *1200C*, 58–65.

Touzani, K., & Sclafani, A. (2005). Critical role of amygdala in flavor but not taste preference learning in rats. *Eur. J. Neurosc.*, *22*(7), 1767–1774.

Vankov, A., Herve-Minville, A., & Sara, S. J. (1995). Response to novelty and its rapid habituation in locus coeruleus neurons of the freely exploring rat. *Eur. J. Neurosci., 7*(6), 1180–1187.

van Oers, H. J., de Kloet, E. R., Li, C., & Levine, S. (1998). The ontogeny of glucocorticoid negative feedback: Influence of maternal deprivation. *Endocrinology*, *139*(6), 2838–2846.

Van Oers, H., de Kloet, E. R., Whelan, T., & Levine, S. (1998). Maternal deprivation effect on the infant's neural stress markers is reversed by tactile stimulation and feeding but not by suppressing corticosterone. *Neuroscience*, *18*, 10171–10179.

Varendi, H., Porter, R. H., & Winberg, J. (1996). Attractiveness of amniotic fluid odor: Evidence of prenatal olfactory learning? *Acta Paediatr*., *85*(10), 1223–1227.

Walker, C., Sapolsky, R., Meaney, M., Vale, W., & Rivier, C. (1986). Increased pituitary sensitivity to glucocorticoid feedback during the stress nonresponsive period in the neonatal rat. *Endocrinology*, *119*(4), 1816–1821.

Weldon, D. A., Travis, M. L., & Kennedy, D. A. (1991). Posttraining D1 receptor blockade impairs odor conditioning in neonatal rats. *Behav. Neurosci.*, *105*(3), 450–458.

Wiedenmayer, C. P., & Barr, G. A. (2001). Developmental changes in c-fos expression to an age-specific social stressor in infant rats. *Behav. Brain Res.*, *126*(1–2), 147–157.

Wilson, D. A., Best, A. R., & Sullivan, R. M. (2004). Plasticity in the olfactory system: Lessons for the neurobiology of memory. *Neuroscientist*, *10*(6), 513–524.

Wilson, D. A., & Stevenson, R. J. (2003). Olfactory perceptual learning: The critical role of memory in odor discrimination. *Neurosci. Biobehav. Rev.*, *27*(4), 307–328.

Wilson, D. A., & Sullivan, R. M. (1994). Neurobiology of associative learning in the neonate: Early olfactory learning. *Behav. Neural Biol.*, *61*(1), 1–18.

Wilson, D. A., Sullivan, R. M., & Leon, M. (1987). Single-unit analysis of postnatal olfactory learning: Modified olfactory bulb output response patterns to learned attractive odors. *J. Neurosci.*, *7*(10), 3154–3162.

Winzer-Serhan, U. H., Raymon, H. K., Broide, R. S., Chen, Y., & Leslie, F. M. (1997). Expression of alpha 2 adrenoceptors during rat brain development I: Alpha 2A messenger RNA expression. *Neuroscience*, *76*(1), 241–260.

Woo, C. C., Coopersmith, R., & Leon, M. (1987). Localized changes in olfactory bulb morphology associated with early olfactory learning. *J. Comp. Neurol.*, *263*(1), 113–125.

Woo, C., & Leon, M. (1988). Sensitive period for neural and behavioral responses to learned odors. *Dev. Brain Res.*, *36*, 309–313.

Yuan, Q., Harley, C. W., Darby-King, A., Neve, R. L., & McLean, J. H. (2003). Early odor preference learning in the rat: Bidirectional effects of camp response element-binding protein (CREB) and mutant CREB support a causal role for phosphorylated CREB. *J. Neurosci.*, *23*(11), 4760–4765.

Yuan, Q., Harley, C. W., McLean, J. H., & Knopfel, T. (2002). Optical imaging of odor preference memory in the rat olfactory bulb. *J. Neurophysiol.*, *87*(6), 3156–3159.

Zeanah, C. H., Keyes, A., & Settles, L. (2003). Attachment relationship experiences and childhood psychopathology. *Ann. NY Acad. Sci.*, *1008*, 22–30.

Zhang, J. J., Okutani, F., Inoure, S., & Kaba, H. (2003). Activation of the cyclic AMP response element-binding protein signaling pathway in the olfactory bulb is required for the acquisition of olfactory aversive learning in young rats. *Neuroscience*, *117*(3), 707–713.

61 Emotional Reaction and Action: From Threat Processing to Goal-Directed Behavior

JOSEPH E. LEDOUX, DANIELA SCHILLER, AND CHRISTOPHER CAIN

ABSTRACT Fear was traditionally studied by using instrumental aversive responses, such as avoidance. This work on instrumental actions failed to lead to a clear understanding of the underlying fear circuitry. Over the past several decades, research on Pavlovian fear reactions has elucidated the circuits that mediate fear. Armed with this information, we return to a consideration of fear-based instrumental actions. While both reactions and actions depend on the amygdala, somewhat different circuits are involved. Fear reactions require connections from the lateral to the central nucleus and from there to the brain stem, while fear-based actions (or at least some such actions) involve connections from the lateral to the basal amygdala and from there to forebrain targets (possibly the striatum). Elucidating the neural mechanisms underlying interactions between Pavlovian and instrumental aversive learning will enhance our understanding of how the brain shifts from passive reactions to actions in the face of danger. This knowledge might aid in understanding how to break the vicious cycle of pathological avoidance in anxiety disorders and could also lead to better coping strategies and other therapeutic interventions.

In 1996, a bomb exploded in Olympic Village in Atlanta. As soon as the explosion occurred, everyone in the crowd was frozen in fear. A few seconds later, they began to run away. This scene, captured on video, illustrates two fundamental ways in which people respond in emotional situations. First we react, then we act (LeDoux, 1996a; LeDoux & Gorman, 2001). Reactions are inflexible responses that are automatically elicited by the stimulus, while actions are instrumental responses that are emitted. Reactions are inevitable consequences that have been programmed by evolution or individual experience, while actions are cognitively controlled responses that are flexibly selected at the moment to achieve a goal.

Research on the neural basis of emotion in animal models has traditionally focused on how emotional stimuli come to elicit fear reactions. Much has been learned over the past several decades, especially about how the brain acquires and controls reactions related to fear-arousing stimuli. Less is known about how emotional actions are acquired and controlled.

Because pathological states involving fear often include the performance of instrumental responses that are maladaptive, this is an important topic to understand. For example, a characteristic feature of pathological anxiety is the escape and avoidance of fear-arousing situations. These can be effective strategies in the short run, since they reduce exposure to situations in which fear arousing stimuli occur and prevent threat escalation. Avoidance can also be effective in the long run, as long as it does not interfere with normal daily life. However, avoidance becomes maladaptive when routine activities are disrupted by excessive or inappropriate avoidance.

In this chapter, we give an overview of the relation between reaction and action in the context of fear. However, we will also consider positive or appetitive emotional states, as research in this area has provided both insights into, and challenges to, work on fear.

Behavioral distinctions between reaction and action

To clarify the distinction between emotional reaction and action requires that we consider these in more detail. To do this, we will put these into the context of a more general taxonomy of behavior.

TAXONOMY OF BEHAVIOR Many behaviors that people and other animals perform fall into one of four categories: reflex, reaction, action, and habit. (For other discussions of this topic, see Balleine & Dickinson, 1998; Cardinal, Parkinson, Hall, & Everitt, 2002; Lang & Davis, 2006; H. H. Yin & Knowlton, 2006.)

A reflex is a stimulus-evoked response that usually involves a single muscle or a limited group of muscles. A puff of air to the eye, for example, elicits a closure of that eye, while painful stimulation of the foot elicits withdrawal of that foot.

JOSEPH E. LEDOUX, DANIELA SCHILLER, and CHRISTOPHER CAIN New York University, Center for Neural Science, New York, New York

A reaction is similar to a reflex in that it is elicited by a specific stimulus, but in contrast to a reflex, it usually involves the whole organism (or at least multiple muscle groups) rather than an isolated muscle group. Freezing behavior, the fearful cessation of movement, is an example of such an organismic response involving muscles throughout the body. Reactions are similar to what ethologists call *species-typical behaviors* or *fixed action patterns* (Bolles, 1970; Lorenz & Tinbergen, 1938; Tinbergen, 1951).

Reflexes and reactions are hard-wired responses that are naturally elicited by certain innate stimuli. In other words, with reflexes and reactions, there is a genetically programmed relationship between certain stimuli and the response. Nevertheless, both reflexes and reactions can come under the control of novel events through associative learning. Rodents, for example, are naturally afraid of cats and foxes. A rat that has been born in a laboratory as part of a family that has been removed from these predators for many generations will still express defensive reactions such as freezing behavior in their presence or when exposed to odors from their hair, urine, or feces (C. D. Blanchard & Blanchard, 1972; R. J. Blanchard & Blanchard, 1969b; Rosen, 2004). However, stimuli associated with predators or pain can, through associative learning, come to elicit freezing reactions (C. D. Blanchard & Blanchard, 1972; R. J. Blanchard & Blanchard, 1969b; Bolles, 1972; Fanselow, 1980). As we will discuss below, novel stimuli come to elicit reactions through Pavlovian conditioning. Such reactions are thus called *learned fear responses*. However, it is important to point out that it is not the response that is learned through Pavlovian conditioning. The response is innate, and learning mainly changes the stimuli that have the capacity to elicit the innate response.

Actions, like reactions, occur at the level of the organism but are not automatically elicited by stimuli. Instead, they are emitted in the presence of certain stimuli that direct behavior toward goals. For this reason, actions are said to be purposive or goal-directed and involve motivation (Balleine & Dickinson, 1998; Niv, Joel, & Dayan, 2006). They are instrumental responses—responses that are instrumental in attaining goals. Running away from the source of a bomb is an example of a behavior that is performed to attain a goal, as is approach toward stimuli related to food or sex. These responses are usually based on past learning and/or information stored in memory that is used to make a decision about what to do. Further distinguishing an action and a reaction is the fact that actions are flexible rather than fixed responses (Ikemoto & Panksepp, 1999). One's natural inclination after freezing momentarily to an explosion might be to run away from the location where the explosion occurred, but if a loved one were closer to the explosion, then the response would probably be to run toward the explosion. Thus a reaction, being hard-wired, is always expressed in the same way in an individual and across individuals in a species. An action, on the other hand, is an arbitrary response that is performed because of its relationship to the goal. You may run, crawl, or swim to escape from a bomb, depending on the situation, but you will likely react first with startle and freezing.

When an instrumental response is performed repeatedly in the presence of certain stimuli, it can become strongly connected to those stimuli and occur inflexibly in the presence of such stimuli. When this situation exists, a stimulus-response (S-R) habit is said to have developed. S-R habits, or just habits, are like reflexes and reactions in that they are elicited by stimuli. However, in contrast to habits and reactions, they are not hard-wired innate responses. Instead, they are based on the transformation of flexible, learned instrumental responses into inflexible responses that have lost their relationship to goal attainment. Habits are often discussed in the context of bad habits, such as nail biting, smoking, overeating, and the like. But not all habits are bad habits. Habits are adaptive and useful when they simplify your life and allow you to perform routine activities faster without having to devote brain resources to them. Habits become pathological when they control behavior in maladaptive ways; for example, a patient who habitually stays home for fear of having a panic attack if she goes outside is stuck in a maladaptive pattern of avoidance behavior. Habits are closely related to skills. Bike riding involves deliberate instrumental learning reinforced by successful movements until the actions become habitual and the skill is acquired. Habits and skills can refer to cognitive as well as behavioral responses.

Historical Note About Research Methodology in the Study of Fear

Having placed reactions and actions in a broader behavioral context, we will now focus on reactions and actions in the context of fear-arousing or threatening situations. However, before we turn to contemporary research on the reactions and actions and the neural systems that are involved, it will be useful to review the history of this field briefly.

A learned fear reaction is a Pavlovian conditioned fear response. Pavlovian fear conditioning is a procedure in which an emotionally neutral conditioned stimulus (CS), such as a tone, is paired with an aversive unconditioned stimulus (US), typically electric shock. Following pairing, the CS acquires the capacity to elicit freezing behavior (R. J. Blanchard & Blanchard, 1969a; Bolles & Fanselow, 1980). Freezing is often said to be a conditioned response (CR). However, freezing is itself not learned. As was noted above, what is learned is an association between the CS and US that allows the CS to elicit a response that it did not elicit previously. This is why Pavlovian conditioning is typically described as a form of stimulus-stimulus (S-S) learning in which the CS acquires emotional potency by its relation to the US. In addition to behavioral responses such as freezing,

physiological changes inside the body also come under the control of the CS, such as changes in autonomic and endocrine activity (Kapp, Frysinger, Gallagher, & Haselton, 1979; LeDoux, Sakaguchi, & Reis, 1983; Schneiderman, Francis, Sampson, & Schwaber, 1974). These are important in providing physiological support for the behavioral responses and in general are part of the integrated organismic response to danger.

Fear conditioning was studied by Pavlov (1927), who called it *defense conditioning*. It was then used by Watson (1929) in his famous study of Little Albert. However, studies of learning in psychology emphasized Thorndike's (1898) instrumental conditioning procedure over Pavlov's approach through the middle of the 20th century and beyond. The behaviorists dominated psychology during this time and believed that instrumental learning was the key to understanding complex human behaviors (Hull, 1943; Skinner, 1938). As a result, a form of instrumental learning called *avoidance conditioning* emerged as the primary behavioral approach to study learned fear (Miller, 1948, 1951; Mowrer, 1947; Mowrer & Lamoreaux, 1946).

In avoidance conditioning studies, animals learn to perform a response (active avoidance) or withhold a response (passive avoidance) in order to avoid harm (typically electric shock). We will be focusing on active avoidance, which we will refer to as *avoidance conditioning* throughout. Some examples of responses that are measured in avoidance studies are shuttling in a runway maze or pressing a lever. Avoidance conditioning is traditionally viewed as a two-stage process. First, stimuli associated with the shock via Pavlovian conditioning are learned about, and then instrumental responses are performed that reduce exposure to those fear-arousing stimuli (Brown & Jacobs, 1949; Kalish, 1954; Levis, 1989; McAllister & McAllister, 1971; Miller, 1948, 1951; Mowrer, 1947; Mowrer & Lamoreaux, 1946; Overmier & Lawry, 1979; Solomon & Wynne, 1954). Therefore the avoidance response could be used to study fear. Pavlovian conditioning was thought to be involved but was not itself measured or studied.

Much research attempted to understand the neural basis of fear through studies of avoidance (Gabriel et al., 1983; Goddard, 1964; Isaacson, 1982; Sarter & Markowitsch, 1985; Weiskrantz, 1956). This research generated complex results that were not easily integrated into a coherent view of the brain mechanisms of avoidance, much less fear. Several factors probably contributed to this, including a failure to separate the Pavlovian and instrumental components of the tasks and the use of a variety of different kinds of avoidance tasks that made different demands on behavior and the brain (Cain & LeDoux, 2007; LeDoux, 1996a).

In the meantime, research on the neural basis of learning and memory, which had also been focused on instrumental behaviors during the behaviorist heyday, turned to Pavlovian conditioning, first in invertebrate studies (Alkon, 1983; Carew, Hawkins, & Kandel, 1983; Dudai, Jan, Byers, Quinn, & Benzer, 1976; Kandel & Spencer, 1968; Walters, Carew, & Kandel, 1979) and then in vertebrates (Cohen, 1974; Thompson, 1976). Research on fear followed suit, and by the early 1980s, the neural mechanisms of fear learning in mammals were much more likely to be studied by using Pavlovian fear conditioning than instrumental avoidance (Davis, 1986; Kapp et al., 1979; Kapp, Pascoe, & Bixler, 1984; LeDoux, Sakaguchi, & Rice, 1984; LeDoux, Thompson, Iadecola, Tucker, & Reis, 1983). Pavlovian conditioning allows the measure of fear directly as CS-elicited reactions rather than indirectly through fear-based instrumental avoidance responses. Because simpler behaviors involve simpler circuits, much progress was made rapidly in understanding the brain mechanisms of fear learning, as is discussed below.

Diverse Functions of an Emotional Stimulus

It is natural to think of a Pavlovian CS as eliciting Pavlovian CRs. However, this in only a part of what goes on. That is, a Pavlovian CS can have effects other than the automatic elicitation of hard-wired, inflexible CRs. Two of these consequences are conditioned reinforcement and conditioned motivation. These are important to consider, as they will help us to revisit the topic of avoidance and instrumental fear.

Conditioned Reinforcement in Avoidance: The Escape from Fear Hypothesis

As has been noted, the avoidance-conditioning literature suggested that a Pavlovian CS contributes to the learning of avoidance responses. How exactly does it do this? A leading view is the escape from fear (EFF) hypothesis (Levis, 1989; McAllister & McAllister, 1971; Miller, 1948; Mowrer, 1947; Rescorla & Solomon, 1967). According to this idea, classical conditioning first establishes the CS as a fear-arousing stimulus, one that elicits fear reactions. As avoidance conditioning proceeds, behaviors that terminate exposure to the fear-arousing CS are reinforced by the reduction in fear that results. Formally, this is called *conditioned negative reinforcement of a stimulus-response association*. *Negative* in this case refers not to the aversive nature of the conditioning but to the fact that the reinforcement comes from termination of the stimulus. A more precise term is *aversive negative conditioned reinforcement*.

Because the Pavlovian and instrumental aspects of avoidance task are intermixed, such tasks cannot be used to directly test the EFF hypothesis. Tasks were therefore developed in which Pavlovian conditioning occurs first and then the CS is presented in a separate chamber. Results showing that behaviors associated with CS termination are learned and repeated support the EFF hypothesis (McAllister & McAllister, 1971). However, the EFF hypothesis of

avoidance has been controversial, due largely to methodological problems with past EFF tasks but also to competing theories of psychological mechanisms underlying avoidance learning (Bolles, 1972; Herrnstein, 1969; Levis, 1989; McAllister & McAllister, 1971; Seligman & Johnston, 1973).

In an effort to address the EFF controversy, we developed a new procedure that controls for many of the factors raised in past critiques (Cain & LeDoux, 2007) (figure 61.1). In our new task, rats learned to rear on their hind legs to terminate a fear-arousing CS. This learning is long-lasting (24 hours) and response-specific (no increase in other nonreinforced behaviors). Interestingly, successful EFF learning also resulted in a transition from passive freezing reactions to escape actions; rats that learned the EFF response showed no spontaneous recovery of freezing following the extinguishing CS presentations that were used for EFF training. Importantly, expression of EFF learning was also controlled by the CS, since animals that went through EFF training did not respond differently than yoked controls until the CS was presented. Thus our data lead us to conclude that instrumental responses can be reinforced by the fear-reducing effects of CS termination.

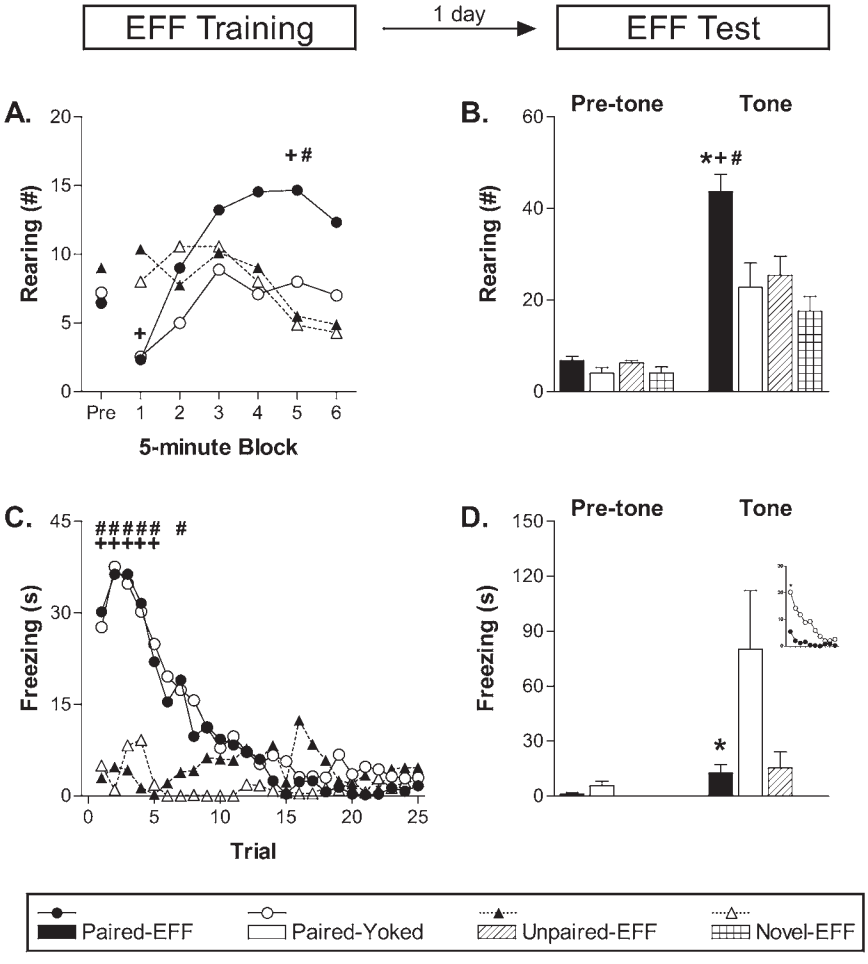


FIGURE 61.1 Escape from fear learning represents instrumental learning that is motivated by fear and reinforced by fear reduction. We designed a new EFF task that controls for factors that made past results inconclusive about the role of fear reduction in EFF learning (Cain & LeDoux, 2007). One day after Pavlovian tone-shock pairings (Paired), tone-shock unpairings (Unpaired), or no training (Novel), rats were presented with 25 tone-alone presentations in a new context (EFF training, *left*). For EFF rats, rearing during a tone presentation led to its immediate termination (response-reinforcement pairing). Yoked rats received identical tone presentations independent of their behavior. One day after EFF training, rats were presented with a single, continuous 10-minute tone presentation to assess long-term EFF memory (EFF test, *right*). Rearing and freezing were assessed during both phases. Paired-EFF rats showed a twofold increase in the EFF escape response (rearing) during the training and testing session compared to Paired-Yoked rats (*A* and *B*). Unpaired-EFF and Novel-EFF rats had no fear of the CS and did not acquire the EFF response (enhanced rearing). Successful acquisition of this active escape response was also associated with less passive freezing to the tone (*D*; inset = minute-by-minute freezing during the EFF test). Further analysis demonstrated that EFF learning was response-specific and performance was motivated by fear (no difference in rearing in the absence of the CS; data not shown). Figure adapted with permission from *Journal of Experimental Psychology: Animal Behavior Processes* (Cain & LeDoux, 2007).

Note that fear-arousing CSs can also serve as positive reinforcers, or punishers. Thus behaviors performed when an aversive CS is presented are less likely to be performed in the future. In this case, the reinforcement is "positive" because learning depends on the delivery of the aversive CS. This contrasts with the role of CS termination in reinforcing escape/avoidance responses. Conditioned punishment, which involves the suppression of response performance by a stimulus, may also be a contributing factor in avoidance and EFF learning (Bolles, 1972). Conditioned punishment is especially relevant to passive avoidance and will not be a focus of this chapter.

Conditioned Motivation In EFF learning (and presumably in avoidance conditioning), the CS functions as a conditioned reinforcer that establishes the acquisition of a new instrumental action. However, once learned, the CS can also function as an incentive that contributes to the performance of the instrumental response.

A conditioned incentive is a stimulus that enhances or suppresses the performance of an existing (previously learned) instrumental action. This incentive-based function of a CS is referred to as *conditioned motivation*. While conditioned motivation probably occurs during the performance of previously conditioned EFF or avoidance responses, as was noted, the preferred test for studying conditioned motivation is something called *Pavlovian-to-instrumental transfer* (PIT). In such tasks, a separately conditioned CS is presented while an animal performs an instrumental behavior.

Most research on PIT and its neural basis has involved appetitive tasks. In these tasks, a previously conditioned CS (say, a tone paired with food) is presented while the subjects perform appetitive instrumental responses (such as food-motivated lever pressing) (Balleine & Killcross, 2006; Blundell, Hall, & Killcross, 2001; Corbit & Balleine, 2005; Everitt, Cardinal, Parkinson, & Robbins, 2003; Holland & Gallagher, 2003; Talmi, Seymour, Dayan, & Dolan, 2008). The typical result is that the CS enhances the performance of the instrumental response, indicating an increase in the motivation to perform the response. The PIT assay is the preferred measure of conditioned motivation because the critical test does not rely on new learning. Independent Pavlovian and instrumental associations are established prior to the test in separate sessions, allowing an uncontaminated assessment of CS-elicited motivation.

The main aversive PIT task in common use is conditioned suppression of appetitive instrumental behavior (Bouton & Bolles, 1980; Estes & Skinner, 1941; Hunt & Brady, 1955). However, this task differs significantly from appetitive PIT tasks. An aversive PIT test that is comparable to appetitive PIT would involve an enhancement of aversive instrumental responding by an aversive CS rather than suppression of appetitive instrumental responding. Such tasks have in fact been used (Brackbill & Overmier, 1979; Grossen & Bolles, 1968; LoLordo, 1967), although virtually no information regarding the neural mechanisms that mediate such aversive PIT exists to date.

Reaction Versus Action: A Summary Behavioral studies have shown that a Pavlovian CS elicits Pavlovian CRs (reactions) but can also influence instrumental actions in at least two ways. One of these is by serving as a conditioned reinforcer that mediates the learning of a new instrumental action, and the other is by serving as a conditioned incentive that affects the motivation to perform a previously learned instrumental action. These are not completely independent functions, since once conditioned reinforcement is used to acquire an instrumental response, the same CS functions as an incentive that motivates the performance of the response. Next we consider what is known about the neural basis of these three functions of a CS, focusing first on reactions.

Neural basis of fear reactions

Research over the past two decades has clearly pointed to the amygdala as a key site where CS-US associations are formed during Pavlovian fear conditioning as well as a necessary site for the later expression of fear reactions elicited by the CS. Below, we will review the role of the amygdala in the fear-conditioning circuitry.

Organization of the Amygdala The amygdala was first recognized as a distinct brain region in the early 19th century (Burdach, 1819–1822). The name, derived from the Greek, was meant to denote an almond-like shape structure in the medial temporal lobe. Like most brain regions, the amygdala is not a single mass but is composed of distinct subareas or nuclei (figure 61.2).

One long-standing idea is that the amygdala consists of an evolutionarily primitive division associated with the olfactory system (cortical, medial, and central nuclei) and an evolutionarily newer division associated with the neocortex (lateral, basal, and accessory basal nuclei) (Johnston, 1923). The areas of the older division are sometimes grouped as the corticomedial region (cortical and medial nuclei) and sometimes as the centromedial region (the central and medial nuclei), while the newer structures related to the neocortex are often referred to as the basolateral region (figure 61.3). The almond-shaped structure that originally defined the amygdala involved the basolateral region rather than the whole structure that is now considered to be the amygdala (Swanson & Petrovich, 1998).

In recent years, there have been a number of attempts to reclassify the amygdala and its relationship to other areas. For example, Heimer (2003) has argued for the concept of an extended amygdala. In this view, the central and medial

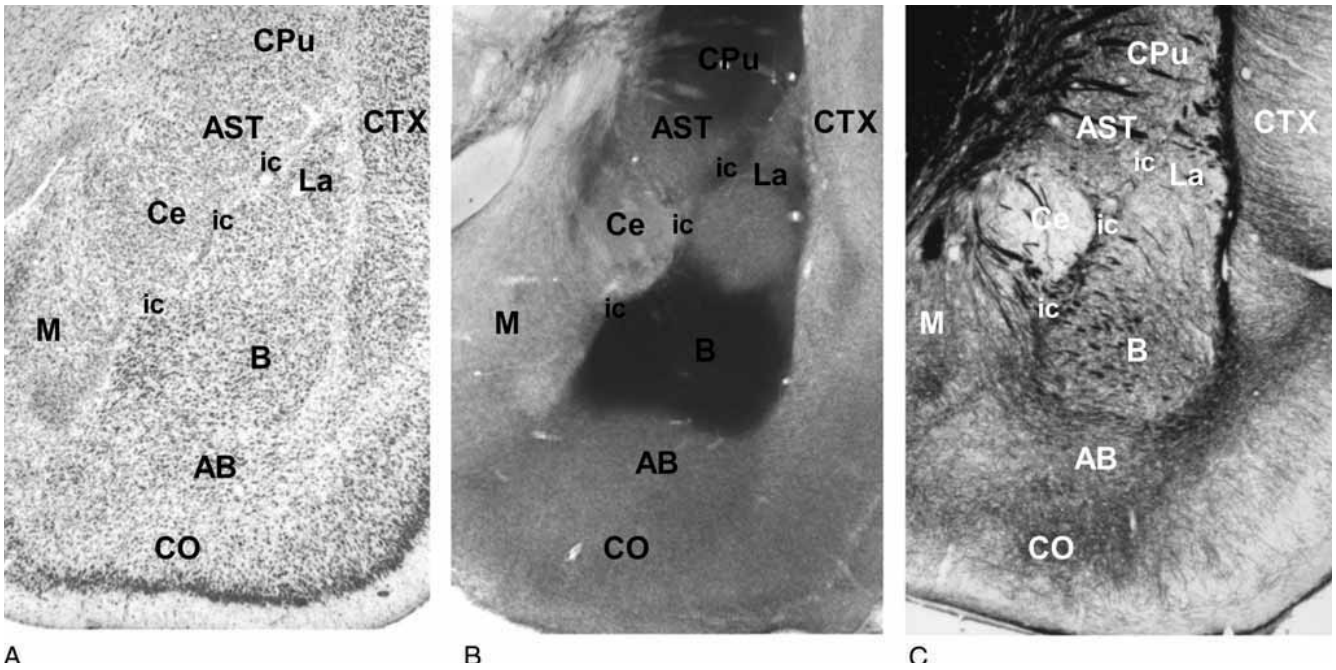


FIGURE 61.2 The rat amygdala. The amygdala of mammals, including humans, consists of at least 12 distinct nuclei. Different staining methods show some of the major nuclei from different perspectives. (*A*) Nissl cell body stain. (*B*) Acetylcholinesterase stain. (*C*) Silver fiber stain. Abbreviations: Amygdala areas: AB, accessory basal; B, basal nucleus; Ce, central nucleus; CO, cortical nucleus; ic, intercalated cells; La: lateral nucleus; M, medial nucleus. Non-amygdala areas: AST, amygdalo-striatal transition area: CPu, caudate putamen; CTX, cortex. (See color plate 75.)

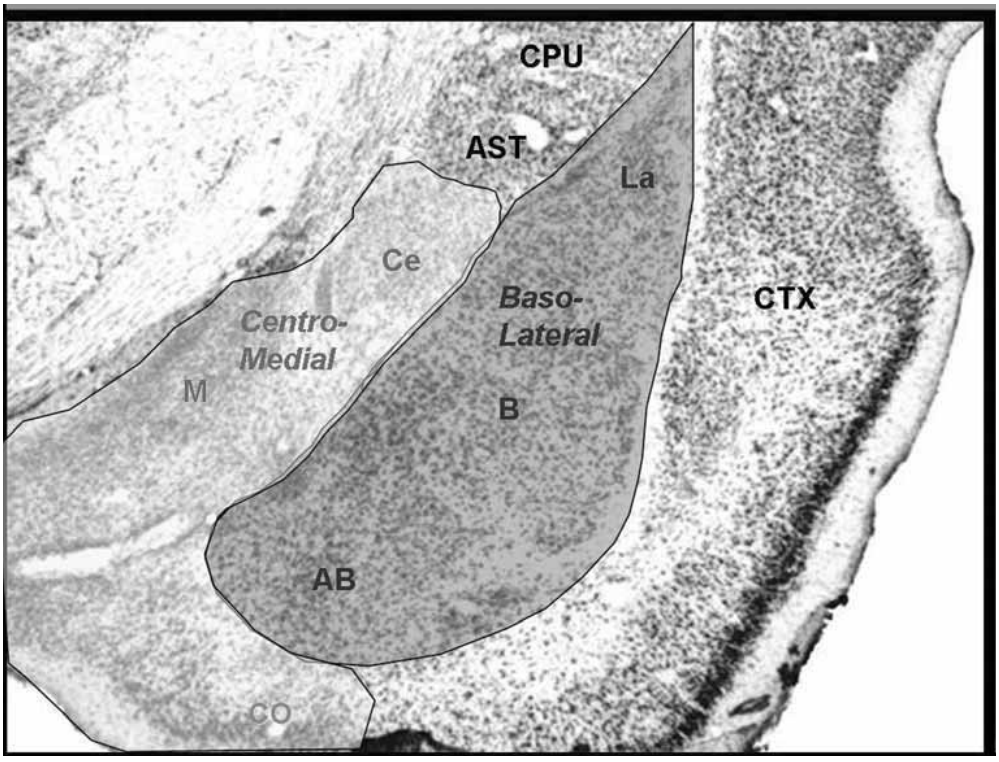


FIGURE 61.3 Groupings of amygdala nuclei. The various nuclei of the amygdala are often partitioned into an evolutionarily old division (the centromedial or corticomedial region) and an evolutionarily newer division (the basolateral region or basolateral complex). While these divisions have some value in understanding the phylogenetic and ontogenetic origins of the amygdala, they do not represent meaningful function divisions, since functions are mediated by cells within much more localized regions, especially subnuclei and even subdivisions of subnuclei. Abbreviations: AB, accessory basal; AST, amygdalo-striatal transition area; B: basal nucleus; Ce, central nucleus; CPu, caudate putamen; CTX, cortex; La, lateral nucleus; M, medial nucleus. (See color plate 76.)

amygdala are continuous (anatomically and neurochemcially) with the lateral and medial divisions of the bed nucleus of the stria terminalis and should be considered a structural unit with functional significance, especially for psychopathology. Swanson and Petrovich (1998) proposed a more radical idea, arguing that "the amygdala," whether extended or not, does not exist as a structural unit. Instead, they argue that the amygdala consists of regions that belong to other regions or systems of the brain and that the designation "the amygdala" is not necessary. For example, in this scheme, the lateral amygdala and basal amygdala are viewed as nuclear extensions of the neocortex (rather than simply as amygdala regions related to the neocortex), the central and medial amygdala are ventral extensions of the striatum, and the cortical nucleus is part of the olfactory system. While this scheme has some merit, the present review focuses on the organization and function of nuclei and subnuclei that are traditionally said to be part of the amygdala, as these perform their functions regardless of whether the amygdala itself exists.

It is easy to be confused by the terminology that is used to describe the amygdala nuclei, as different sets of terms are used. This problem is especially acute with regards to the basolateral region of the amygdala. As was noted, the basolateral region consists of the lateral, basal, and accessory basal nuclei. However, in another terminological scheme, the basal and accessory basal nuclei are called the *basolateral* and *basomedial nuclei*, respectively. The use of the term *basolateral* to refer both to a specific nucleus (the basal or basolateral nucleus) and to the larger region that includes the lateral, basal, and accessory basal nuclei (the basolateral region) is the source of some difficulty, since authors do not always clearly identify whether they are referring to the nucleus or the region. Further, some studies use the term *basolateral complex* (BLA) to refer to the lateral and basal nuclei (and usually not the accessory basal).

Each of the nuclei of the amygdala can be further partitioned into subnuclei (Pitkänen, 2000; Pitkänen, Pikkarainen, Nurminen, & Ylinen, 1997). For example, the lateral nucleus has three major divisions: dorsal, ventrolateral, and medial (figure 61.4). Further division is also possible. The dorsal subdivsion has a superior and an inferior region. The central nucleus, on the other hand, has lateral, capsular, and medial divisions. These subnuclear partitions of the lateral and central nuclei have turned out to have important functional significance.

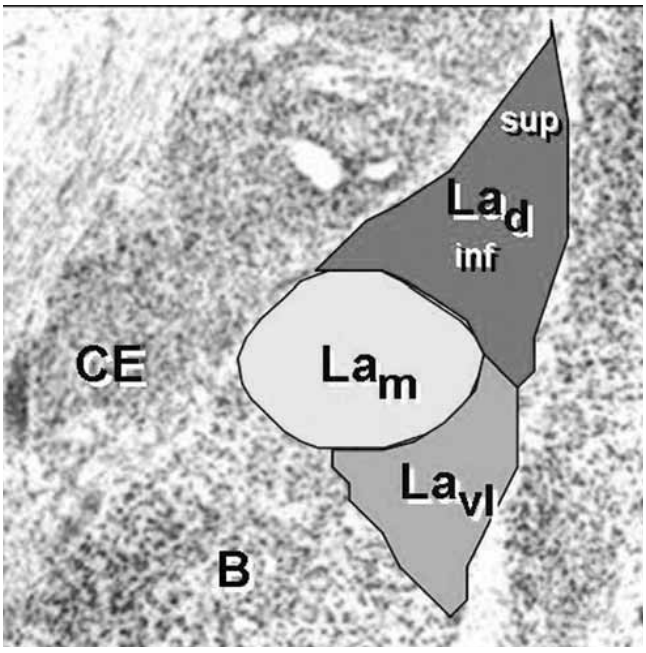


FIGURE 61.4 Subdivisions of the lateral nucleus of the amygdala. The lateral nucleus of the amygdala has three major subdivision: dorsal (LA_d), ventrolateral (LA_{vl}), and medial (LA_m). Each of these has additional partitions. The dorsal subnucleus, for example, contains a superior (sup) and inferior (inf) region. Cells in the superior region have been implicated in the acquisition of fear conditioning, and cells have been implicated in the inferior region in long-term memory storage (see text). Abbreviations: B, basal nucleus; CE, central nucleus. (See color plate 77.)

STANDARD MODEL OF CONDITIONED FEAR REACTIONS: SERIAL PROCESSING WITHIN THE AMYGDALA Two areas of the amygdala are generally considered to be especially important for the acquisition and expression of Pavlovian fear conditioning (figure 61.5). The lateral nucleus (LA) receives and integrates the CS and US and later processes the CS in the control of fear reactions. The central amygdala (CE), on the other hand, is especially involved in the expression of fear reactions. The basal nucleus (B) and the intercalated cell masses (ICT) also contribute.

The standard model evolved from a series of studies that first implicated the central amygdala (CE) starting in the late 1970s (Hitchcock & Davis, 1986; Kapp et al., 1979, 1984; LeDoux, Iwata, Cicchetti, & Reis, 1988). Given its connections to hypothalamic and brain stem areas that control species-typical behaviors and autonomic and endocrine responses related to fear, CE was viewed as important for the expression of fear-related CRs. Studies showing that the firing rate of cells in CE increased following CS-US pairing (Pascoe & Kapp, 1985) suggested that CE may also be a key site of plasticity in the formation of the CS-US association.

In the 1990s, emphasis began to shift to the lateral amygdala (LA) when it was shown that sensory inputs from the CS and US pathways mainly terminate in LA rather than in CE (Bordi & LeDoux, 1992; Clugnet & LeDoux, 1990; LeDoux, Sakaguchi, Iwata, & Reis, 1986; LeDoux et al., 1984), that damage to LA disrupts fear conditioning (LeDoux, Cicchetti, Xagoraris, & Romanski, 1990), and that CS and US inputs converge on single cells in LA

(Romanski, LeDoux, Clugnet, & Bordi, 1993). These studies suggested that CS-US convergence does not take place in the entire LA but only in its dorsal subnucleus (Romanski et al., 1993). Subsequent studies found that CS-US pairing during fear conditioning leads to the enhancement of short-latency neural responses elicited by a CS in the dorsal LA (Quirk, Armony, & LeDoux, 1995). Subsequent studies confirmed the importance of the dorsal LA (Collins & Pare, 2000; Hobin, Goosens, & Maren, 2003; Maren, 2000; Maren & Quirk, 2004; Repa et al., 2001) and further suggested that there may be different cell groups within the dorsal LA involved in the initial learning and long-term storage of plasticity (Repa et al., 2001).

Thus, CS-US convergence in LA came to be viewed as inducing synaptic plasticity. When the CS later occurs, it is transmitted to LA, where the potentiated synapses drive CE to express fear reactions via output connection to the brainstem (Davis, Walker, & Lee, 1997; Fanselow & LeDoux, 1999; LeDoux, 1996b, 2000; Maren, 2001; Maren & Fanselow, 1996). This standard model of fear conditioning is sometimes called the serial model because it assumes that inputs come into LA, that LA connects with CE, and that CE connects with regions that control CRs.

The idea that LA is a key site of plasticity was given additional weight by findings showing that long-term synaptic potentiation (LTP) could be induced in LA (Clugnet & LeDoux, 1990) or BLA (Chapman, Kairiss, Keenan, & Brown, 1990) by patterns of electrical stimulation. Particularly important was the fact that LTP could be induced in CS pathways to LA in vivo (Clugnet & LeDoux, 1990; Rogan & LeDoux, 1995) and in vitro (Weisskopf, Bauer, & LeDoux, 1999). Further, fear conditioning itself was shown to produce LTP-like changes in CS processing in LA in vivo (Rosenkranz & Grace, 2002; Rogan, Staubli, & LeDoux, 1997) and in vitro (McKernan & Shinnick-Gallagher, 1997).

Additional studies showed that fear conditioning and LTP in LA depend on similar molecular mechanisms. Initial studies focused on NMDA receptors, since these were known to be involved in synaptic plasticity in the hippocampus (Collingridge & Bliss, 1995; Cotman, Monaghan, & Ganong, 1988). Thus blockade of NMDA receptors in the LA disrupted fear conditioning (Fanselow & Kim, 1994; Maren, Aharonov, Stote, & Fanselow, 1996; Miserendino, Sananes, Melia, & Davis, 1990; Rodrigues, Schafe, & LeDoux, 2001), and similar manipulations also disrupted LTP in LA (Bauer, Schafe, & LeDoux, 2002). Alteration of gene expression related to AMPA receptors in LA also alters LTP and fear conditioning (Rumpel, LeDoux, Zador, & Malinow, 2005). Moreover, fear conditioning and LTP depend on similar second messenger and macromolecular events in LA (Dityatev & Bolshakov, 2005; Fanselow & Poulos, 2005; LeDoux, 2000; Maren, 2001; Pape & Stork, 2003; Rodrigues, Schafe, & LeDoux, 2004; Sah, Faber, Lopez De Armentia, & Power, 2003; Walker & Davis, 2002). Studies of genetically altered mice also revealed parallels between fear conditioning and LTP (Matynia, Kushner, & Silva, 2002; Mayford, Abel, & Kandel, 1995; Wang, Hu, & Tsien, 2006), and some of the mechanisms overlap with findings from invertebrates (Bailey, Giustetto, Huang, Hawkins, & Kandel, 2000; Dubnau, Chiang, & Tully, 2003; Lechner & Byrne, 1998; Roberts & Glanzman, 2003; J. C. Yin & Tully, 1996).

Given that LA is the main sensory input region of the amygdala and CE is the output link to the brain stem, the standard model requires that connections exist between LA and CE. The LA connects with the CE via several direct and indirect routes within the amygdala (Pare, Quirk, & LeDoux, 2004; Pitkänen, Savander, & LeDoux, 1997; Royer, Martina, & Pare, 1999). Because it is generally assumed that the medial nucleus of CE (CEm) connects with brain stem areas that control CRs, a key issue is the extent to which LA connects with CEm (Pare et al., 2004). While direct connections to CEm seem weak at best, the LA connects with the lateral CE, which has connections to CEm (Petrovich & Swanson, 1997), though these latter connections are somewhat sparse. The LA also connects with the intercalate cell group (ITC), which form a chain of connections that lead to CEm (Royer et al., 1999). The LA also connects to the CEm via the B and accessory basal nuclei (Pitkänen et al., 1997). The multiplicity of connections between LA and CE suggest why lesions of a structure such as B may have effects on fear conditioning in some studies (Anglada-Figueroa & Quirk, 2005; Goosens & Maren, 2001) but not others (Amorapanth, LeDoux, & Nader, 2000), depending possibly on details of the training paradigm or possibly on the extent of damage in B. Regardless, paths of connectivity within the amygdala clearly link LA to CE, including CEm.

Recently, there has been a revival of interest in the possibility that the CE may be a site of plasticity as well as an output link between LA and areas that control CRs (Pare et al., 2004; Samson, Duvarci, & Pare, 2005; Wilensky, Schafe, Kristensen, & LeDoux, 2006). Several studies have thus shown that plasticity occurs in CE in vitro (Fu & Shinnick-Gallagher, 2005; Samson & Pare, 2005). Most germane is the fact that CE undergoes plasticity in vivo during aversive conditioning, as determined by single-unit recordings (Pascoe & Kapp, 1985; Rorick-Kehn & Steinmetz, 2005). Our recent findings showing that functional inactivation or protein synthesis inhibition in CE prevents memory formation during Pavlovian conditioning suggests that plasticity in CE plays a key and essential role (Wilensky et al., 2006). However, the earliest latency of unit CRs and the number of trials required to acquire unit CRs in CE (Pascoe & Kapp, 1985) are both greater than those in LA (Goosens, Hobin, & Maren, 2003; Maren, 2000; Maren & Quirk, 2004; Quirk, Armony, & LeDoux, 1997; Quirk et al., 1995; Repa et al.,

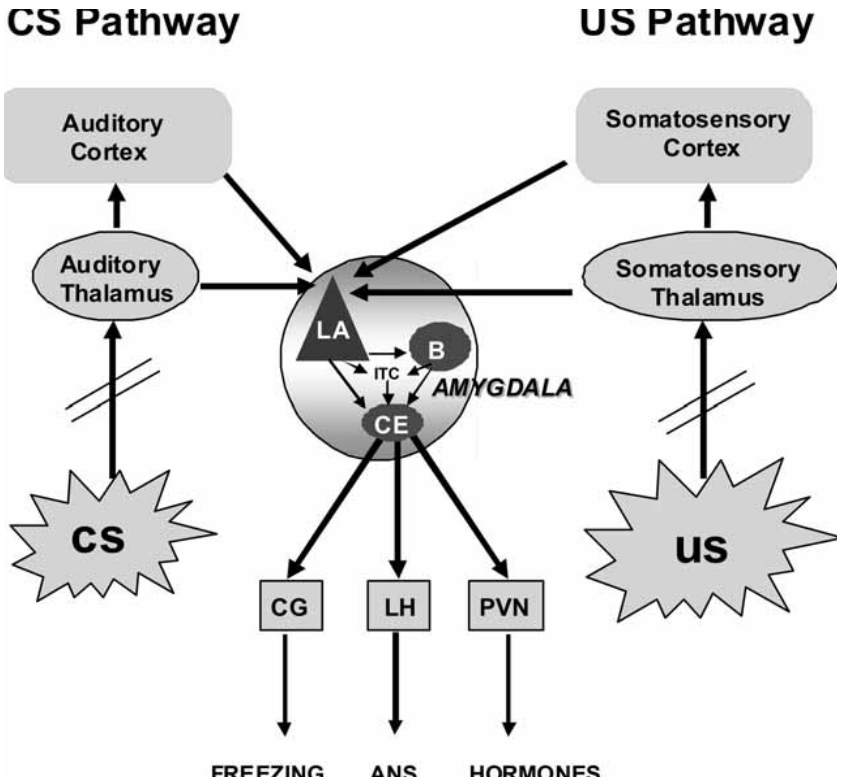


FIGURE 61.5 Auditory fear conditioning pathways. The auditory conditioned stimulus (CS) and somatosensory (pain) unconditioned stimulus (US) converge in the lateral amygdala (LA). The LA receives inputs from each system via both thalamic and cortical inputs. CS-US convergence induces synaptic plasticity in LA such that after conditioning the CS flows through the LA to activate the central amygdala (CE) via intra-amygdala connections. CE in turn controls the expression of behavioral (e.g., freezing), autonomic (ANS), and endocrine responses that are components of the fear reaction. Abbreviations: B, basal amygdala; CG, central gray; LH, lateral hypothalamus; ITC, intercalated cells of the amygdala; PVN, paraventricular nucleus of the hypothalamus.

2001). These data are consistent with the notion that LA and CE are both sites of plasticity but that CE plasticity depends on LA plasticity.

The fact that LA is normally required for fear conditioning does not rule out the possibility that under some circumstances, fear conditioning can occur in animals with lesions of LA and B. Indeed, fear conditioning can occur in animals with lesions of LA and B when extensive training is given, especially in contextual conditioning but also in cued conditioning in some cases (Hall, Thomas, & Everitt, 2000; Killcross, Robbins, & Everitt, 1997; Lee, Dickinson, & Everitt, 2005; Maren, 1998, 1999).

A CHALLENGE TO THE STANDARD MODEL: PARALLEL PROCESSING WITHIN THE AMYGDALA In spite of the extensive evidence in support of the standard model, this view has been challenged. In these challenges, LA and B are usually considered together as an undifferentiated structure that is referred to as the BLA. It is argued that some aspects of fear conditioning can be mediated by the BLA independent of CE and that other aspects can be mediated by CE independent of BLA. Because BLA and CE are proposed to receive CS and US inputs separately and function independently, this is called the *parallel model* (Balleine & Killcross, 2006; Cardinal et al., 2002; Killcross et al., 1997; Killcross & Blundell, 2002).

First of all, proponents of the parallel view argue that fear conditioning can occur when BLA is damaged (Balleine & Killcross, 2006). While this is true, essentially all of the evidence for this has come from studies in which overtraining is used (Hall, Parkinson, Connor, Dickinson, & Everitt, 2000; Killcross et al., 1997; Lee et al., 2005; Maren, 1998, 1999). Most studies that have contributed to the serial model have involved training with a small number of trials (fewer than 10). With overtraining, weak intra-amygdala pathways, weak direct sensory inputs to CE, or more complex circuitous pathways involving other brain regions may undergo synaptic plasticity, potentiation of which may allow the pathways to be utilized in ways that are not possible with few training trials. Indeed, recent studies show that CE is necessary for learning in overtrained animals (Zimmerman, Rabinak, McLachlan, & Maren, 2007). Thus, while findings from overtraining are important and interesting, they likely involve different circuits than those involved in the rapid form of fear conditioning that occurs in natural situations in which organisms do not always have the opportunity to

"practice" learning what to fear over many trials. BLA, and especially LA, is essential when standard, naturalistic training is involved.

Appetitive conditioning research is strongly influenced by the notion that CSs can enter into associations with different aspects of the US (sensory versus affective properties) and that these separate associations generate different response types (consummatory versus preparatory), an idea that was first formalized by Konorski (1967). These ideas were particularly important for the development of the parallel model of amygdala function, which is largely based on appetitive findings. Consummatory responses are those that are engaged when the subject is in direct contact with the US, such as chewing and licking in the case of food. Preparatory responses for food include approach and autonomic nervous system changes, responses that are appropriate when food is in the vicinity. Note that both consummatory and preparatory responses can include reactions and actions. Appetitive research on brain mechanisms of conditioning has therefore focused on identifying structures that mediate these separate associations and separate response classes. In the process, this research has evaluated both reactions and actions. This is in contrast to aversive research, which has primarily focused on fear reactions, at least in the last few decades. Aversive studies, less influenced by the Konorskian distinction between consummatory and preparatory responses, have generally assumed that there is one basic association between the CS and all aspects of the US. And this basic association can be used in different ways to generate different basic behaviors, including both reactions and actions. Thus, embedded within the serial versus parallel amygdala processing debate are fundamental questions regarding the nature of the CS-US association(s) and basic response classes.

In an abstract sense, it would be logical for the brain to process appetitive and aversive stimuli in similar ways. However, this might not be the case. Despite the superficial similarity of the behavioral designs that are used for appetitive and aversive conditioning, distinct kinds of stimuli serve as the US (food or water versus shock), the associations involved are fundamentally different (tone + food or water versus tone + shock), the conditioning procedures require many more training trials for appetitive conditioning relative to aversive conditioning, and the responses (approach versus freezing) likely involve different output circuits of the amygdala and possibly different intra-amygdala circuits.

The serial model of fear conditioning has been based on studies that measure conditioned reactions elicited by the CS. In contrast, the evidence for the parallel model has mostly been based on studies in which instrumental responses are used as an indirect measure of conditioned fear. In particular, much of the evidence for the parallel model has relied on the use of Pavlovian to instrumental transfer (PIT) tests in which a previously conditioned CS alters the performance of a previously trained instrumental response. These studies are extremely important, as they have provided important clues about how to pursue the role of the amygdala in fear-related instrumental actions as an assessment of conditioned motivation. But their value in understanding how associations are acquired and stored during fear conditioning is less clear. Below, when we discuss the neural basis of PIT, we will discuss these findings and their applicability (or lack thereof) to fear conditioning.

In sum, the evidence that is being used to defend the parallel model does not justify the rejection of decades of research on fear conditioning. This will become more apparent later, after we discuss the neural basis of fear actions, especially the contribution of conditioned motivation, as studied by PIT tasks.

Neural basis of fear actions

Earlier, we mentioned two ways in which a Pavlovian fear-arousing CS influences instrumental responses: by serving as a conditioned reinforcer and by serving as a conditioned motivator (incentive). These are studied by using EFF and PIT tasks. In this section, we will therefore review research on the neural basis of EFF and PIT.

Amygdala Contributions to Conditioned Reinforcement: Studies of EFF A leading hypothesis about avoidance is that the instrumental response is reinforced by termination of stimuli (CSs) associated with the aversive US. That is, the avoidance response is learned and initially performed to remove the CS. Hence, CS termination serves as a conditioned negative reinforcer, a stimulus that increases the likelihood of responses that eliminates this same stimulus. This is usually studied by using EFF tasks. In spite of its potential to explain avoidance, little research has studied the role of the brain in EFF tasks.

The only study that explored the brain mechanisms of EFF was performed by our laboratory (Amorapanth et al., 2000) (figure 61.6). We found that pretraining lesions of LA disrupted both Pavlovian conditioning and EFF learning. CE lesions disrupted Pavlovian freezing responses but not instrumental EFF responses. And B lesions disrupted EFF learning but not freezing. Specifically, our results suggested that CS information flow through the amygdala begins in LA, where the Pavlovian CS-US association is stored, and then this association is used to generate a conditioned reinforcing signal (negative or positive) either in B or in its output targets. The double dissociation that was observed for CE and B led to the proposal that LA is critical for both aversive reaction and action learning because the Pavlovian CS-US association was learned and stored in LA, and this association is critical for both types of responses. Serial processing from LA to CE mediated aversive reaction learning, and serial processing

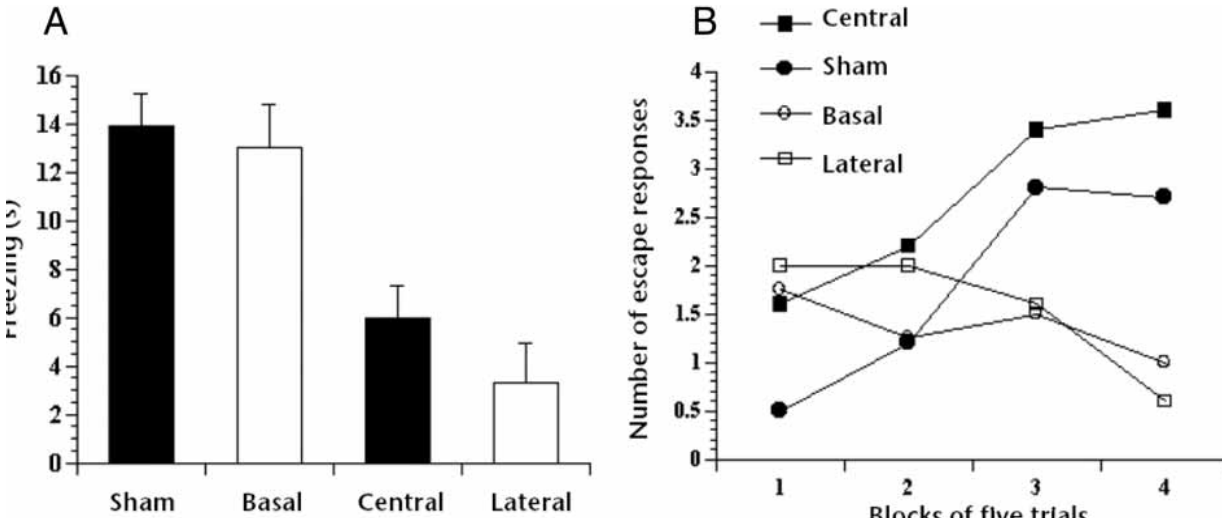


FIGURE 61.6 Escape from fear (EFF) learning depends on the lateral and basal, but not central, amygdala. Prior to behavioral training, rats received bilateral electrolytic lesions of LA, CE, or B. Rats were first subjected to Pavlovian fear conditioning and then to EFF training, using chamber crossing as the escape response. LA and CE lesions disrupted performance of a passive fear reaction to the CS (*A*, freezing). LA and B lesions disrupted performance of an active EFF response (*B*, chamber crossing). These data suggest that LA is necessary for establishing the CS as a conditioned incentive. This information is then relayed to CE to initiate passive Pavlovian reactions and to B for active escape responding. Note that this study was done with an older EFF task prior to the development of the task illustrated in figure 61.1. (Reproduced from Amorapanth, LeDoux, & Nader (2000) with permission of *Nature Neuroscience*.)

from LA to B mediated aversive action learning (figure 61.7). Because the tasks that we and others before employed had several shortcomings, we developed a new task (Cain & LeDoux, 2007). This task will be especially useful in drawing firm conclusions about the role of amygdala areas in EFF learning and hence conditioned reinforcement.

Another study that is relevant was performed by Killcross and colleagues (1997). They trained rats on a concurrent Pavlovian conditioning and conditioned punishment task and assessed the effects of CE versus BLA lesions. In this task, a previously conditioned CS punishes and thereby weakens a previously established, appetitively motivated instrumental response. In the same sessions, Pavlovian conditioning was assessed by conditioned suppression of appetitive bar pressing, using a separate bar from that used for the punishment assessment. Their results suggested that BLA lesions, but not CE lesions, interfered with conditioned punishment, which matches well with the results of Amorapanth and colleagues. But they also found that conditioned suppression, the Pavlovian measure, was impaired by CE lesions but not by BLA lesions. This second finding is at odds with many studies on conditioned reactions supporting the serial model. However, this was a complex task involving significant overtraining, and the relevance of the results to simpler tasks that do not involve overtraining should be viewed cautiously (Lee et al., 2005). Specifically, typical fear-conditioning studies use fewer than 10 training trials (as few as one in many studies), whereas Killcross and colleagues used approximately 120 trials. Also, the instrumental task involved conditioned punishment rather than conditioned negative reinforcement. Further, they did not distinguish between LA and B. Nevertheless, the general conclusion regarding fear-based action is the same as that from our study: LA and B are essential for this influence of Pavlovian CS on instrumental learning.

These conclusions from aversive conditioning are also consistent with appetitive conditioning results that indicate that the BLA but not the CE contributes to conditioned reinforcement (Burns, Everitt, & Robbins, 1999). The appetitive work suggests further that connections from the BLA to the ventral striatum allow the conditioned reinforcer, formed in the BLA, to mediate the acquisition of the instrumental action. Extra-amygdala circuits will be discussed below.

AMYGDALA CONTRIBUTIONS TO CONDITIONED MOTIVATION: STUDIES OF PIT Excluding conditioned suppression studies, research on brain mechanisms of PIT mainly involves appetitive conditioning tasks in which a previously conditioned CS is used as a conditioned incentive and its effects on instrumental behavior are assessed (Balleine & Killcross, 2006; Corbit & Balleine, 2005; Everitt et al., 2003; Hall et al., 2001; Holland & Gallagher, 2003; Talmi et al., 2008). This work indicates that the amygdala is critical for appetitive PIT. Further, as will be described below, different amygdala nuclei appear to make distinct contributions to different forms of appetitive PIT.

Prior to PIT studies, early appetitive conditioning research identified separate contributions of CE and BLA to general affective responses and US-specific responses. Damage to

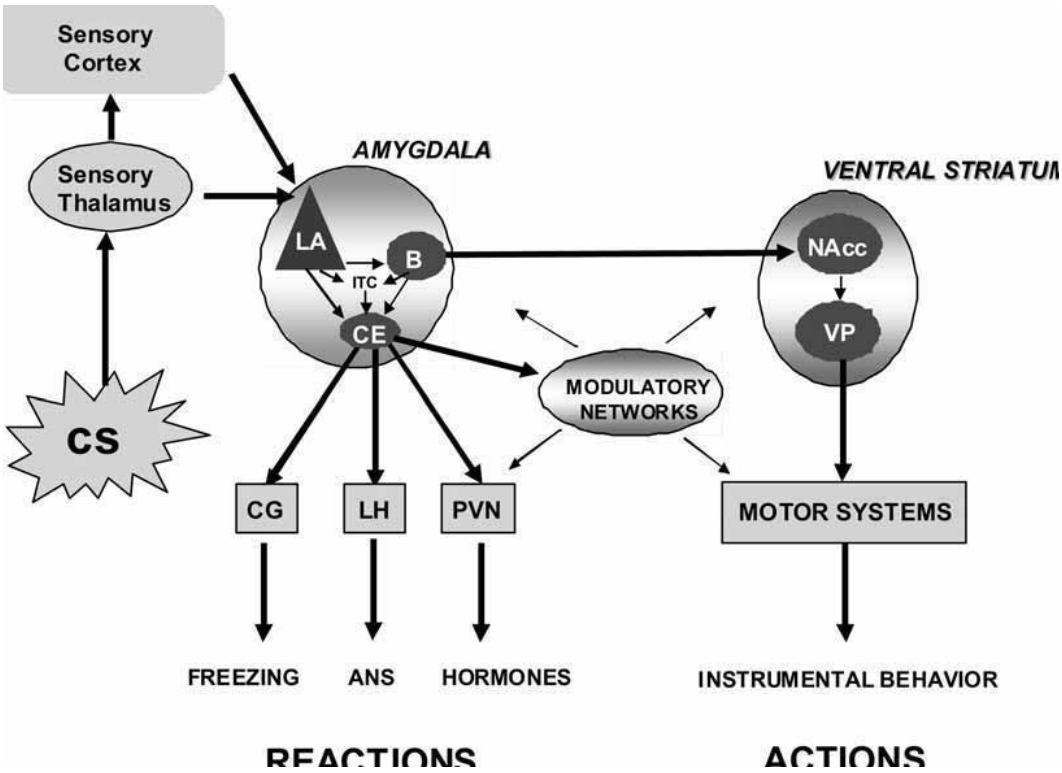


FIGURE 61.7 Schematic representation of circuits proposed to mediate fear-elicited emotional reactions and fear-motivated emotional actions. During fear conditioning the CS-US association is formed and stored in the lateral amygdala (LA) (see figure 61.5). After fear conditioning, the CS, processed in the sensory thalamus and cortex, drives activity in the LA and then CE. Activity in CE, in turn, leads to the expression of passive fear reactions (such as freezing behavior), activation of brain stem neuromodulatory systems that release amine neurotransmitters throughout the brain, and activation of autonomic nervous system (ANS) and neuroendocrine responses. The CS also functions as an incentive and a conditioned reinforcer by way of information flow from LA to B. Projections from B to nucleus accumbens (NAcc) may allow the use of CS information in the control of instrumental actions. Specifically, the NAcc processing of incentive information is proposed to invigorate and guide active behavior via projections to the ventral pallidum (VP) and downstream motor systems, with the aid of neuromodulators (especially dopamine arriving from ventral tegmental area). Note that the amygdala aspects of the model are derived from work on amygdala-dependent fear conditioning and EFF learning (Amorapanth et al., 2000), while downstream portions of the model are borrowed from work in appetitive conditioning (e.g., Ikemoto & Panksepp, 1999; Kalivas & Nakamura, 1999; Cardinal et al., 2002). Once EFF is well-learned, there is a hypothetical inhibition of passive fear actions. Additional abbreviations: PAG, periaqueductal gray; LH, lateral hypothalamus; PVN, paraventricular nucleus; LA, lateral amygdala; B, basal amygdala; CE, central amygdala.

CE, but not to BLA, impaired the "preparatory" conditioned approach (Parkinson, Robbins, & Everitt, 2000) and orienting (Gallagher, Graham, & Holland, 1990). An opposite result was found with US-specific "consummatory" responses: Lesions of BLA, but not of CE, interfered with US devaluation effects (Hatfield, Han, Conley, Gallegher, & Holland, 1996) and potentiation of feeding by an appetitive CS (Gallagher & Holland, 1992; Holland & Petrovich, 2005; Holland, Petrovich, & Gallagher, 2002). Later studies pursued this dissociation, using appetitive PIT tasks (Balleine & Killcross, 2006; Corbit & Balleine, 2005; Everitt et al., 2003; Hall et al., 2001; Holland & Gallagher, 2003), in which Pavlovian CSs enhanced instrumental responding for food. Positive PIT (enhancement of responding) was observed when the USs in Pavlovian and instrumental conditioning matched (US-specific PIT) and when they were both appetitive but distinct (general PIT). BLA lesions disrupted US-specific PIT but had no effect on US-general PIT, while CE lesions disrupted general but not specific PIT. Contradictory results in earlier studies (Everitt et al., 2003; Hall et al., 2001; Holland & Gallagher, 2003) have been attributed to the failure to distinguish between specific and general PIT (Balleine & Killcross, 2006).

Balleine and Killcross (2006) and others (Cardinal et al., 2002; Killcross & Blundell, 2002) argue that the model inspired by appetitive conditioning results should also apply to aversive motivation, based in large part on the effects of amygdala lesions on conditioned suppression. In conditioned suppression, an aversive CS decreases food-motivated lever pressing. This is viewed as a form of general PIT. The finding that suppression depends on CE and not on BLA (Killcross et al., 1997) is thus used to argue that aversive

general PIT depends on CE and not on BLA. Such data were the basis of the parallel model of amygdala processing in fear conditioning discussed above.

There are several reasons not to accept this conclusion without further study. First, conditioned suppression is not methodologically comparable to appetitive PIT. A comparable task would be one in which an aversive CS enhances an aversive instrumental response rather one in which an aversive CS suppresses an appetitive instrumental response. Second, the conclusion that conditioned suppression depends on CE but not on BLA is mainly supported by a study that used overtraining (Killcross et al., 1997), which, as already noted, is problematic. Further, a number of other studies have found that BLA lesions affect conditioned suppression, especially in the absence of overtraining (Cousens & Otto, 1998; Lee et al., 2005; Schiller & Weiner, 2004; Selden, Everitt, Jarrard, & Robbins, 1991). To account for this involvement of BLA in suppression when only a few training trials are used, it has been argued that suppression in such situations is due to freezing, which is deemed a US-specific response and not "true suppression," which is considered a general affective response (Balleine & Killcross, 2006). This view is challenged by our finding that PAG-lesioned animals exhibit suppression but not freezing, even when only a few training trails are used (Amorapanth, Nader, & LeDoux, 1999). Since freezing is eliminated by the PAG lesion, the suppression that occurs must count as "true suppression" in spite of the fact that only a few training trials were used. Thus, freezing is not necessarily the explanation for why BLA lesions disrupt suppression. If suppression does reflect a form of general motivation based on emotional arousal, then it might be the case that CE is involved in general PIT but as an output of LA and B rather than a parallel and independent pathway.

To directly test the role of different amygdala areas in specific and general motivational functions, US-specific and US-general PIT tasks are needed. These have not been developed and used in aversive studies of brain function to date. However, it is likely that such tasks could be developed, since studies of avoidance conditioning have found that an aversive CS can enhance avoidance responding (Brackbill & Overmier, 1979; Ehrman & Overmier, 1976; Grossen & Bolles, 1968; LoLordo, 1967).

The PIT studies thus return us to the question of serial versus parallel processing in the amygdala. The appetitive studies point to BLA and CE as separately and independently mediating different forms of associative learning, leading to different responses. However, in aversive conditioning, the results suggest that LA and B are likely to be involved in both US-specific and general PIT. The CE may also be involved in aversive general PIT, but if so, this will occur via LA (and possibly B) rather than via the formation of a CS-US association that is independent of LA and B.

As has been noted already, we should not expect the brain to process appetitive and aversive stimuli in similar ways, given how different the USs and the responses involved are. Thus rather than overturning decades of research, the appetitive results provide hypotheses to be tested and methods for testing them in fear conditioning. Only after such work has been done should we draw the conclusion that amygdala circuits mediating fear conditioning are the ones that have been identified for appetitive conditioning rather than the ones that have been supported by decades of research.

Extra-Amygdala Areas Involved in Conditioned Reinforcement and Conditioned Motivation

The evidence reviewed above suggests that the basal amygdala (B) is likely to play a key role in at least some aspects of aversive conditioned reinforcement and motivation. Anatomical output connections of B are thus probable target areas that might function as postamygdala processing links (Amaral & Insausti, 1992; Gabbott, Warner, & Busby, 2006; Johnson, Aylward, Hussain, & Totterdell, 1994; Kelley, Domesick, & Nauta, 1982; McDonald, 1998; Pitkanen et al., 2000). It is also possible that some aspects of conditioned motivation or reinforcement might involve connections from B to CE. This will not be considered further here. Instead, we will focus on four extra-amygdala targets of B: nucleus accumbens (NAcc), ventromedial prefrontal cortex (vmPFC), anterior cingulate cortex (ACC), and orbitofrontal cortex (OFC).

Nucleus accumbens NAcc is important for instrumental learning in response to environmental stimuli in appetitive tasks (Kelley, 2004) and may be required for expression of this memory, once it has been learned (Ikemoto & Panksepp, 1999; Nicola, 2007). Specifically, NAcc contributes to motivated action learning (Cardinal & Everitt, 2004; Ikemoto & Panksepp, 1999; Kelley, 2004; Koob, 1996; Salamone, Correa, Mingote, & Weber, 2003), including action learning about conditioned reinforcers (de Borchgrave, Rawlins, Dickinson, & Balleine, 2002; Kelley & Delfs, 1991; Robbins, Giardini, Jones, Reading, & Sahakian, 1990) (see figure 61.7).

Anatomically, NAcc is situated at the neural crossroads of emotion and movement (Graybiel, 1976; Mogenson, Jones, & Yim, 1980). Importantly, NAcc neuronal processing appears to contribute to the expression of aversively motivated instrumental learning as intra-NAcc dopamine receptor antagonists disrupt performance of signaled avoidance learning (Wadenberg, Ericson, Magnusson, & Ahlenius, 1990). In addition, performance of conditioned avoidance behavior is reduced by systemic catecholamine synthesis inhibition and rescued by intra-NAcc injections of dopamine (Jackson, Ahlenius, Anden, & Engel, 1977). Recent studies have also shown that NAcc dopamine is

elevated in response to aversive CS presentations (Salamone, Correa, Farrar, & Mingote, 2007). Thus we predict that pre-EFF lesions of the NAcc will impair EFF learning and performance. The above studies also suggest that NAcc may play a role in aversive conditioned motivation, and therefore we predict that posttraining lesions, or disconnections from B, may also impair aversive PIT. NAcc core and shell may make unique contributions, since NAcc core lesions seem to have a greater effect on appetitive conditioned motivation and conditioned reinforcement than shell lesions do (Hall et al., 2001; Parkinson, Olmstead, Burns, Robbins, & Everitt, 1999).

Ventromedial prefrontal cortex vmPFC consists of the infralimbic and prelimbic areas. Studies of interest have not always distinguished between the two regions. Although the literature on vmPFC contributions to conditioned reinforcement or conditioned motivation is sparse, some recent studies have implicated this region in expression of goal-directed actions (Balleine & Dickinson, 1998; Coutureau & Killcross, 2003). In addition, it has been suggested that the connection between B and vmPFC represents a functional link between incentive value and instrumental contingencies (Cardinal et al., 2002). Of particular relevance, however, is the well-documented role of the vmPFC in behavioral inhibition, especially inhibition of passive fear reactions such as freezing. The vmPFC, especially the infralimbic cortex, plays a prominent role in extinction (suppression) of Pavlovian fear reactions (Quirk, Garcia, & Gonzalez-Lima, 2006; Sotres-Bayon, Cain, & LeDoux, 2006). We hypothesize that vmPFC plays some role in enabling action learning like avoidance or EFF and predict that pretraining lesions, or disconnections from B, of infralimbic cortex will impair the acquisition of avoidance and EFF.

Anterior cingulate cortex (ACC) The ACC, which lies dorsal to vmPFC, has been the subject of a good deal of research in emotional learning tasks, both appetitive and aversive. The literature is a bit confusing, but several findings suggest that the ACC may be an interesting target, especially in EFF. For instance, ACC lesions impair signaled active avoidance learning in rabbits (Gabriel, Vogt, Kubota, Poremba, & Kang, 1991), and CS-evoked neural responses in ACC develop early in avoidance training (Gabriel, 1990). It has been suggested that BLA and ACC coordinate to add specificity to instrumental actions that depend on Pavlovian CSs (Everitt et al., 2003). Available data suggest that ACC may be particularly important for instrumental actions that depend on Pavlovian CSs (conditioned reinforcement; reviewed in Cardinal et al., 2002). Thus we predict that lesions of ACC or ACC disconnection from B will impair EFF learning. ACC might not be necessary for PIT, at least in appetitive conditioning, since ACC lesions do not affect the enhancement of appetitive instrumental responding by a CS previously paired with food (Cardinal et al., 2003).

Orbitofrontal cortex OFC has been implicated in various behaviors involving the integration of incentive value with instrumental actions (Cardinal et al., 2002; Rolls, 2004; Schoenbaum, Gottfried, Murray, & Ramus, 2007). For instance, rats, monkeys, and humans with OFC damage show risk assessment impairments in instrumental choice tasks and an inability to flexibly alter behavior as contingencies change (Bechara, Damasio, & Damasio, 2000; Dias, Robbins, & Robbins, 1996; Gallagher & Schoenbaum, 1999). These deficits seem to be dependent on interactions between OFC and B, since disconnection lesions of these two structures create similar impairments in incentive processing. On the basis of these findings, we predict that OFC damage and OFC disconnection from B will impair both EFF learning and PIT performance.

Summary and conclusions

Animals, including humans, can respond in many ways to aversive stimulation. These responses can be broadly separated into four categories: reflexes, reactions, instrumental actions, and habits. Although fear research initially focused on instrumental avoidance behavior, this strategy was largely abandoned in favor of simple Pavlovian conditioned reactions. This switch in focus to a simpler procedure probably occurred because it facilitated investigations of brain mechanisms and because instrumental avoidance was believed to depend, at least partly, on prior Pavlovian conditioning. This new tack proved to be successful, and several decades' worth of intense research has led to a fairly comprehensive understanding of how the brain mediates the learning, storage, and performance of conditioned fear reactions. Briefly, the amygdala is critical for all three, LA being important for learning and storage and CE being important especially for expression of fear reactions. A recent modification of the standard serial model suggests that B is important for the reinforcement and motivation of instrumental actions, using the CS-US association stored in LA. Thus the standard serial model of fear conditioning assumes that LA is critical for generating the CS-US association, and CE and B mediate different consequences of this learning: reactions versus actions. However, much new research is needed to examine this hypothesis in more depth.

The switch in focus to basic fear reactions also had its costs, as our understanding of aversive action learning such as avoidance and EFF has lagged far behind. Such responses represent a large class of both normal and pathological fear behavior and should be studied in more detail to complete

our understanding of fear-related behavior. It should also be noted that fear actions likely depend on, or at least interact with, prior Pavlovian conditioning. Thus, the foundation provided by decades of intense Pavlovian conditioning research is likely to speed our understanding of aversive instrumental actions.

Research on appetitive conditioning has progressed along a slightly different trajectory. Most notably, appetitive research has benefitted from the Konorskian separation of responses into US-specific "consummatory" and general affective "preparatory" responses. Psychological assays have been honed to tap into these processes, and studies of neural mechanisms have focused on this distinction. The results of this research led to the development of the parallel model of amygdala function, in which CE mediates associations between the CS and general affective properties of the US, and generates responses appropriate for this association. BLA is said to mediate associations between the CS and sensory properties of the US and generates responses appropriate to this association. Note that each association type can generate both actions and reactions, so the distinction here is different from that emphasized by fear research. Importantly, CE and BLA processing are thought to function separately and independently, a concept that is also at odds with the standard fear model. And of course, the models differ in their conception of the association(s) that are formed. In contrast to the parallel model, the serial model assumes that one CS-US association type is formed in LA, between the CS and all properties of the US.

Appetitive conditioning research had progressed greatly in the past few decades and can inform new studies in the aversive field regarding conditioned reinforcement and conditioned motivation of instrumental actions. However, we argue that it is much too soon to abandon the serial model of amygdala-dependent fear conditioning, which is supported by a significant amount of past research. There are just too many methodological (and other) differences between aversive and appetitive studies that can explain the differences between the models. And there is the very real possibility that the amgydala functions one way for appetitive behaviors and another way for aversive behaviors. We argue that the appropriate path forward is to design new experiments to directly test the involvement of various amgydala regions in specific aspects of aversive instrumental action. More data are needed to resolve this apparent discrepancy. Appetitive studies will be invaluable in aiding the direction of this research.

In conclusion, much has been learned about how the brain acquires and controls reactions triggered by fear-arousing stimuli. However, much less is known about the neural mechanisms allowing the acquisition and control of fear-motivated actions. Given the importance of actions to normal coping and pathological states, it is imperative that we begin to unravel these complex processes. Pathological fear states often involve instrumental responses that were adaptive but under certain conditions became maladaptive and inappropriate. Elucidating the neural mechanisms that underlie interactions between Pavlovian and instrumental aversive learning might enhance our understanding of what makes the shift from passive reactions to actions possible in the face of fear. This knowledge may aid our ability to break the vicious cycle of avoidance and lead to better coping strategies and therapeutic interventions.

REFERENCES

Alkon, D. L. (1983). Learning in a marine snail. *Sci. Am.*, *249*, 70–85.

Amaral, D. G., & Insausti, R. (1992). Retrograde transport of d-[3h]-aspartate injected into the monkey amygdaloid complex. *Exp. Brain Res.*, *88*(2), 375–388.

Amorapanth, P., LeDoux, J. E., & Nader, K. (2000). Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nat. Neurosci.*, *3*(1), 74–79.

Amorapanth, P., Nader, K., & LeDoux, J. E. (1999). Lesions of periaqueductal gray dissociate-conditioned freezing from conditioned suppression behavior in rats. *Learn. Memory*, *6*(5), 491–499.

Anglada-Figueroa, D., & Quirk, G. J. (2005). Lesions of the basal amygdala block expression of conditioned fear but not extinction. *J. Neurosci.*, *25*(42), 9680–9685.

Bailey, C. H., Giustetto, M., Huang, Y. Y., Hawkins, R. D., & Kandel, E. R. (2000). Is heterosynaptic modulation essential for stabilizing Hebbian plasticity and memory? *Nat. Rev. Neurosci.*, *1*(1), 11–20.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4–5), 407–419.

Balleine, B. W., & Killcross, S. (2006). Parallel incentive processing: An integrated view of amygdala function. *Trends Neurosci.*, *29*(5), 272–279.

Bauer, E. P., Schafe, G. E., & LeDoux, J. E. (2002). NMDA receptors and l-type voltage-gated calcium channels contribute to long-term potentiation and different components of fear memory formation in the lateral amygdala. *J. Neurosci.*, *22*(12), 5239–5249.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex*, *10*(3), 295–307.

Blanchard, C. D., & Blanchard, R. J. (1972). Innate and conditioned reactions to threat in rats with amygdaloid lesions. *J. Comp. Physiol. Psychol.*, *81*, 281–290.

Blanchard, R. J., & Blanchard, D. C. (1969a). Crouching as an index of fear. *J. Comp. Physiol. Psychol.*, *67*(3), 370–375.

Blanchard, R. J., & Blanchard, D. C. (1969b). Passive and active reactions to fear-eliciting stimuli. *J. Comp. Physiol. Psychol.*, *68*(1), 129–135.

Blundell, P., Hall, G., & Killcross, S. (2001). Lesions of the basolateral amygdala disrupt selective aspects of reinforcer representation in rats. *J. Neurosci.*, *21*(22), 9018–9026.

Bolles, R. C. (1970). Species-specific defense reactions and avoidance learning. *Psychol. Rev.*, *77*, 32–48.

Bolles, R. C. (1972). The avoidance learning problem. In G. H. Bower & J. T. Spencer (Eds.), *The psychology of learning and motivation* (Vol. 6, pp. 97–145). Oxford, UK: Academic Press.

Bolles, R. C., & Fanselow, M. S. (1980). A perceptual-defensive-recuperative model of fear and pain. *Behav. Brain Sci.*, *3*, 291–323.

Bordi, F., & LeDoux, J. (1992). Sensory tuning beyond the sensory system: An initial analysis of auditory response properties of neurons in the lateral amygdaloid nucleus and overlying areas of the striatum. *J. Neurosci.*, *12*(7), 2493–2503.

Bouton, M. E., & Bolles, R. C. (1980). Conditioned fear assessed by freezing and by the suppression of three different baselines. *Anim. Learn. Behav.*, *8*, 429–434.

Brackbill, R. M., & Overmier, J. B. (1979). Aversive CS control of instrumental avoidance as a function of selected parameters and method of Pavlovian conditioning. *Learn. Motiv.*, *10*(3), 229–244.

Brown, J. S., & Jacobs, A. (1949). The role of fear in the motivation and acquisition of responses. *J. Exp. Psychol.*, *39*(6), 747–759.

Burdach, K. F. (1819–1822). *Vom Baue und Leben des Gehirns*. Leipzig: Dyk.

Burns, L. H., Everitt, B. J., & Robbins, T. W. (1999). Effects of excitotoxic lesions of the basolateral amygdala on conditional discrimination learning with primary and conditioned reinforcement. *Behav. Brain Res.*, *100*(1–2), 123–133.

Cain, C. K., & LeDoux, J. E. (2007). Escape from fear: A detailed behavioral analysis of two atypical responses reinforced by CS termination. *J. Exp. Psychol. [Anim. Behav.]*, *33*(4), 451–463.

Cardinal, R. N., & Everitt, B. J. (2004). Neural and psychological mechanisms underlying appetitive learning: Links to drug addiction. *Curr. Opin. Neurobiol.*, *14*(2), 156–162.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.*, *26*(3), 321–352.

Cardinal, R. N., Parkinson, J. A., Marbini, H. D., Toner, A. J., Bussey, T. J., Robbins, T. W., et al. (2003). Role of the anterior cingulate cortex in the control over behavior by Pavlovian conditioned stimuli in rats. *Behav. Neurosci.*, *117*(3), 566–587.

Carew, T. J., Hawkins, R. D., & Kandel, E. R. (1983). Differential classical conditioning of a defensive withdrawal reflex in aplysia californica. *Science*, *219*, 397–400.

Chapman, P. F., Kairiss, E. W., Keenan, C. L., & Brown, T. H. (1990). Long-term synaptic potentiation in the amygdala. *Synapse*, *6*(3), 271–278.

Clugnet, M. C., & LeDoux, J. E. (1990). Synaptic plasticity in fear conditioning circuits: Induction of ltp in the lateral nucleus of the amygdala by stimulation of the medial geniculate body. *J. Neurosci.*, *10*(8), 2818–2824.

Cohen, D. H. (1974). The neural pathways and informational flow mediating a conditioned autonomic response. In L. V. Di Cara (Ed.), *Limbic and autonomic nervous system research* (pp. 223–275). New York: Plenum Press.

Collingridge, G. L., & Bliss, T. V. (1995). Memories of nmda receptors and ltp. *Trends Neurosci.*, *18*(2), 54–56.

Collins, D. R., & Pare, D. (2000). Differential fear conditioning induces reciprocal changes in the sensory responses of lateral amygdala neurons to the cs(+) and cs(−). *Learn. Memory*, *7*(2), 97–103.

Corbit, L. H., & Balleine, B. W. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *J. Neurosci.*, *25*(4), 962–970.

Cotman, C. W., Monaghan, D. T., & Ganong, A. H. (1988). Excitatory amino acid neurotransmission: NMDA receptors and Hebb-type synaptic plasticity. *Annu. Rev. Neurosci.*, *11*, 61–80.

Cousens, G., & Otto, T. (1998). Both pre- and posttraining excitotoxic lesions of the basolateral amygdala abolish the expression of olfactory and contextual fear conditioning. *Behav. Neurosci.*, *112*(5), 1092–1103.

Coutureau, E., & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behav. Brain Res.*, *146*(1–2), 167–174.

Davis, M. (1986). Pharmacological and anatomical analysis of fear conditioning using the fear-potentiated startle paradigm. *Behav. Neurosci.*, *100*(6), 814–824.

Davis, M., Walker, D. L., & Lee, Y. (1997). Amygdala and bed nucleus of the stria terminalis: Differential roles in fear and anxiety measured with the acoustic startle reflex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *352*, 1675–1687.

de Borchgrave, R., Rawlins, J. N., Dickinson, A., & Balleine, B. W. (2002). Effects of cytotoxic nucleus accumbens lesions on instrumental conditioning in rats. *Exp. Brain Res.*, *144*(1), 50–68.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, *380*, 69–72.

Dityatev, A. E., & Bolshakov, V. Y. (2005). Amygdala, long-term potentiation, and fear conditioning. *Neuroscientist*, *11*(1), 75–88.

Dubnau, J., Chiang, A. S., & Tully, T. (2003). Neural substrates of memory: From synapse to system. *J. Neurobiol.*, *54*(1), 238–253.

Dudai, Y., Jan, Y. N., Byers, D., Quinn, W. G., & Benzer, S. (1976). Dunce, a mutant of drosophila deficient in learning. *Proc. Natl. Acad. Sci. USA*, *73*(5), 1684–1688.

Ehrman, R. M., & Overmier, J. B. (1976). Dissimilarity of mechanisms for evocation of escape and avoidance responding in dogs. *Anim. Learn. Behav.*, *4*(3), 347–351.

Estes, W. K., & Skinner, B. F. (1941). Some quantitative properties of anxiety. *J. Exp. Psychol.*, *29*, 390–400.

Everitt, B. J., Cardinal, R. N., Parkinson, J. A., & Robbins, T. W. (2003). Appetitive behavior: Impact of amygdala-dependent mechanisms of emotional learning. *Ann. NY Acad. Sci.*, *985*, 233–250.

Fanselow, M. S. (1980). Conditional and unconditional components of postshock freezing. *Pavlov. J. Biol. Sci.*, *15*, 177–182.

Fanselow, M. S., & Kim, J. J. (1994). Acquisition of contextual Pavlovian fear conditioning is blocked by application of an nmda receptor antagonist d,l-2-amino-5-phosphonovaleric acid to the basolateral amygdala. *Behav. Neurosci.*, *108*, 210–212.

Fanselow, M. S., & LeDoux, J. E. (1999). Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron*, *23*(2), 229–232.

Fanselow, M. S., & Poulos, A. M. (2005). The neuroscience of mammalian associative learning. *Annu. Rev. Psychol.*, *56*, 207–234.

Fu, Y., & Shinnick-Gallagher, P. (2005). Two intra-amygdaloid pathways to the central amygdala exhibit different mechanisms of long-term potentiation. *J. Neurophysiol.*, *93*(5), 3012–3015.

Gabbott, P. L., Warner, T. A., & Busby, S. J. (2006). Amygdala input monosynaptically innervates parvalbumin immunoreactive local circuit neurons in rat medial prefrontal cortex. *Neuroscience*, *139*(3), 1039–1048.

Gabriel, M. (1990). Functions of anterior and posterior cingulate cortex during avoidance learning in rabbits. *Prog. Brain Res.*, *85*, 467–482; discussion 482–463.

Gabriel, M., Lambert, R. W., Foster, K., Orona, E., Sparenborg, S., & Maiorca, R. R. (1983). Anterior thalamic lesions and neuronal activity in the cingulate and retrosplenial cortices during discriminative avoidance behavior in rabbits. *Behav. Neurosci.*, *97*, 675–696.

Gabriel, M., Vogt, B. A., Kubota, Y., Poremba, A., & Kang, E. (1991). Training-stage related neuronal plasticity in limbic thalamus and cingulate cortex during learning: A possible key to mnemonic retrieval. *Behav. Brain Res.*, *46*, 175–185.

Gallagher, M., Graham, P. W., & Holland, P. C. (1990). The amygdala central nucleus and appetitive pavlovian conditioning: Lesions impair one class of conditioned behavior. *J. Neurosci.*, *105*, 1906–1911.

Gallagher, M., & Holland, P. C. (1992). Understanding the function of the central nucleus: Is simple conditioning enough? In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (pp. 307–321). New York: Wiley-Liss.

Gallagher, M., & Schoenbaum, G. (1999). Functions of the amygdala and related forebrain areas in attention and cognition. *Ann. NY Acad. Sci.*, *877*, 397–411.

Goddard, G. (1964). Functions of the amygdala. *Psychol. Rev.*, *62*, 89–109.

Goosens, K. A., Hobin, J. A., & Maren, S. (2003). Auditory-evoked spike firing in the lateral amygdala and Pavlovian fear conditioning: Mnemonic code or fear bias? *Neuron*, *40*(5), 1013–1022.

Goosens, K. A., & Maren, S. (2001). Contextual and auditory fear conditioning are mediated by the lateral, basal, and central amygdaloid nuclei in rats. *Learn. Memory*, *8*(3), 148–155.

Graybiel, A. (1976). *Input-output anatomy of the basal ganglia.* Lecture at the Society for Neuroscience, Toronto, Canada.

Grossen, N., & Bolles, R. (1968). Effects of a classical conditioned "fear signal" and "safety signal" on non-discriminated avoidance behavior. *Psychon. Sci.*, *11*(9), 321–322.

Hall, J., Parkinson, J. A., Connor, T. M., Dickinson, A., & Everitt, B. J. (2001). Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour. *Eur. J. Neurosci.*, *13*(10), 1984–1992.

Hall, J., Thomas, K. L., & Everitt, B. J. (2000). Rapid and selective induction of bdnf expression in the hippocampus during contextual learning. *Nat. Neurosci.*, *3*(6), 533–535.

Hatfield, T., Han, J. S., Conley, M., Gallagher, M., & Holland, P. (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *J. Neurosci.*, *16*(16), 5256–5265.

Heimer, L. (2003). A new anatomical framework for neuropsychiatric disorders and drug abuse. *Am. J. Psychiatry*, *160*(10), 1726–1739.

Herrnstein, R. J. (1969). Method and theory in the study of avoidance. *Psychol. Rev.*, *76*(1), 49–69.

Hitchcock, J., & Davis, M. (1986). Lesions of the amygdala but not of the cerebellum or red nucleus block conditioned fear as measured with the potentiated startle paradigm. *Behav. Neurosci.*, *100*, 11–22.

Hobin, J. A., Goosens, K. A., & Maren, S. (2003). Context-dependent neuronal activity in the lateral amygdala represents fear memories after extinction. *J. Neurosci.*, *23*(23), 8410–8416.

Holland, P. C., & Gallagher, M. (2003). Double dissociation of the effects of lesions of basolateral and central amygdala on conditioned stimulus-potentiated feeding and Pavlovian-instrumental transfer. *Eur. J. Neurosci.*, *17*(8), 1680–1694.

Holland, P. C., & Petrovich, G. D. (2005). A neural systems analysis of the potentiation of feeding by conditioned stimuli. *Physiol. Behav.*, *86*(5), 747–761.

Holland, P. C., Petrovich, G. D., & Gallagher, M. (2002). The effects of amygdala lesions on conditioned stimulus-potentiated eating in rats. *Physiol. Behav.*, *76*(1), 117–129.

Hull, C. L. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

Hunt, H. F., & Brady, J. V. (1955). Some effects of punishment and intercurrent "anxiety" on a simple operant. *J. Comp. Physiol. Psychol.*, *48*, 305–310.

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: A unifying interpretation with special reference to reward-seeking. *Brain Res. Brain Res. Rev.*, *31*(1), 6–41.

Isaacson, R. L. (1982). *The limbic system.* New York: Plenum Press.

Jackson, D. M., Ahlenius, S., Anden, N. E., & Engel, J. (1977). Antagonism by locally applied dopamine into the nucleus accumbens or the corpus striatum of alpha-methyltyrosine-induced disruption of conditioned avoidance behaviour. *J. Neural Transm.*, *41*(4), 231–239.

Johnson, L. R., Aylward, R. L., Hussain, Z., & Totterdell, S. (1994). Input from the amygdala to the rat nucleus accumbens: Its relationship with tyrosine hydroxylase immunoreactivity and identified neurons. *Neuroscience*, *61*(4), 851–865.

Johnston, J. B. (1923). Further contributions to the study of the evolution of the forebrain. *J. Comp. Neurol.*, *35*, 337–481.

Kalivas, P. W., & Nakamura, M. (1999). Neural systems for behavioral activation and reward. *Curr. Opin. Neurobiol.*, *9*, 223–227.

Kalish, H. I. (1954). Strength of fear as a function of the number of acquisition and extinction trials. *J. Exp. Psychol.*, *47*(1), 1–9.

Kalivas, P. W., & Nakamura, M. (1999). Neural systems for behavioral activation and reward. *Curr. Opin. Neurobiol.*, *9*, 223–227.

Kandel, E. R., & Spencer, W. A. (1968). Cellular neurophysiological approaches to the study of learning. *Physiol. Rev.*, *48*, 65–134.

Kapp, B. S., Frysinger, R. C., Gallagher, M., & Haselton, J. R. (1979). Amygdala central nucleus lesions: Effect on heart rate conditioning in the rabbit. *Physiol. Behav.*, *23*(6), 1109–1117.

Kapp, B. S., Pascoe, J. P., & Bixler, M. A. (1984). The amygdala: A neuroanatomical systems approach to its contributions to aversive conditioning. In N. Buttlers & L. R. Squire (Eds.), *Neuropsychology of memory* (pp. 473–488). New York: Guilford.

Kelley, A. E. (2004). Ventral striatal control of appetitive motivation: Role in ingestive behavior and reward-related learning. *Neurosci. Biobehav. Rev.*, *27*(8), 765–776.

Kelley, A. E., & Delfs, J. M. (1991). Dopamine and conditioned reinforcement: I. Differential effects of amphetamine microinjections into striatal subregions. *Psychopharmacology (Berl.)*, *103*(2), 187–196.

Kelley, A. E., Domesick, V. B., & Nauta, W. J. (1982). The amygdalostriatal projection in the rat: An anatomical study by anterograde and retrograde tracing methods. *Neuroscience*, *7*(3), 615–630.

Killcross, S., Robbins, T. W., & Everitt, B. J. (1997). Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. *Nature*, *388*(6640), 377–380.

Killcross, S., & Blundell, P. (2002) Associative representations of emotionally significant outcomes. In S. Moore & Oaksford (Eds.,) *Emotion and cognition: From brain to behaviour* (pp. 35–73). Amsterdam/Philadelphia: John Benjamins.

Konorski, J. (1967). *Integrative activity of the brain.* Chicago: University of Chicago Press.

Koob, G. F. (1996). Hedonic valence, dopamine and motivation. *Mol. Psychiatry*, *1*(3), 186–189.

Lang, P. J., & Davis, M. (2006). Emotion, motivation, and the brain: Reflex foundations in animal and human research. *Progr. Brain Res.*, *156*, 3–29.

Lechner, H. A., & Byrne, J. H. (1998). New perspectives on classical conditioning: A synthesis of Hebbian and non-Hebbian mechanisms. *Neuron*, *20*(3), 355–358.

LeDoux, J. E. (1996a). *The emotional brain.* New York: Simon & Schuster.

LeDoux, J. E. (1996b). Emotional networks and motor control: A fearful view. *Progr. Brain Res.*, *107*, 437–446.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.*, *23*, 155–184.

LeDoux, J. E., Cicchetti, P., Xagoraris, A., & Romanski, L. M. (1990). The lateral amygdaloid nucleus: Sensory interface of the amygdala in fear conditioning. *J. Neurosci.*, *10*(4), 1062–1069.

LeDoux, J. E., & Gorman, J. M. (2001). A call to action: Overcoming anxiety through active coping. *Am. J. Psychiatry*, *158*(12), 1953–1955.

LeDoux, J. E., Iwata, J., Cicchetti, P., & Reis, D. J. (1988). Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *J. Neurosci.*, *8*(7), 2517–2529.

LeDoux, J. E., Sakaguchi, A., Iwata, J., & Reis, D. J. (1986). Interruption of projections from the medial geniculate body to an archi-neostriatal field disrupts the classical conditioning of emotional responses to acoustic stimuli. *Neuroscience*, *17*(3), 615–627.

LeDoux, J. E., Sakaguchi, A., & Reis, D. J. (1983). Strain differences in fear between spontaneously hypertensive and normotensive rats. *Brain Res.*, *277*(1), 137–143.

LeDoux, J. E., Sakaguchi, A., & Reis, D. J. (1984). Subcortical efferent projections of the medial geniculate nucleus mediate emotional responses conditioned to acoustic stimuli. *J. Neurosci.*, *4*(3), 683–698.

LeDoux, J. E., Thompson, M. E., Iadecola, C., Tucker, L. W., & Reis, D. J. (1983). Local cerebral blood flow increases during auditory and emotional processing in the conscious rat. *Science*, *221*(4610), 576–578.

Lee, J. L., Dickinson, A., & Everitt, B. J. (2005). Conditioned suppression and freezing as measures of aversive Pavlovian conditioning: Effects of discrete amygdala lesions and overtraining. *Behav. Brain Res.*, *159*(2), 221–233.

Levis, D. J. (1989). The case for a return to a two-factor theory of avoidance: The failure of non-fear interpretations. In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory* (pp. 227–277). Hillsdale, NJ: Lawrence Erlbaum.

LoLordo, V. M. (1967). Similarity of conditioned fear responses based upon diferent aversive events. *J. Comp. Physiol. Psychol.*, *64*, 154–158.

Lorenz, K. Z., & Tinbergen, N. (1938). Taxis und instinkt-begriffe in der eirollbewegung der graugans. *Z. Tierpsychol.*, *2*, 1–29.

Maren, S. (1998). Overtraining does not mitigate contextual fear conditioning deficits produced by neurotoxic lesions of the basolateral amygdala. *J. Neurosci.*, *18*(8), 3088–3097.

Maren, S. (1999). Neurotoxic basolateral amygdala lesions impair learning and memory but not the performance of conditional fear in rats. *J. Neurosci.*, *19*(19), 8696–8703.

Maren, S. (2000). Auditory fear conditioning increases CS-elicited spike firing in lateral amygdala neurons even after extensive overtraining. *Eur. J. Neurosci.*, *12*(11), 4047–4054.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.*, *24*, 897–931.

Maren, S., Aharonov, G., Stote, D. L., & Fanselow, M. S. (1996). *N*-methyl-D-aspartate receptors in the basolateral amygdala are required for both acquisition and expression of conditional fear in rats. *Behav. Neurosci.*, *110*(6), 1365–1374.

Maren, S., & Fanselow, M. S. (1996). The amygdala and fear conditioning: Has the nut been cracked? *Neuron*, *16*(2), 237–240.

Maren, S., & Quirk, G. J. (2004). Neuronal signalling of fear memory. *Nat. Rev. Neurosci.*, *5*(11), 844–852.

Matynia, A., Kushner, S. A., & Silva, A. J. (2002). Genetic approaches to molecular and cellular cognition: A focus on LTP and learning and memory. *Annu. Rev. Genet.*, *36*, 687–720.

Mayford, M., Abel, T., & Kandel, E. R. (1995). Transgenic approaches to cognition. *Curr. Opin. Neurobiol.*, *5*, 141–148.

McAllister, W. R., & McAllister, D. E. (1971). Behavioral measurement of conditioned fear. In F. R. Brush (Ed.), *Aversive conditioning and learning* (pp. 105–179). New York: Academic Press.

McDonald, A. J. (1998). Cortical pathways to the mammalian amygdala. *Progr. Neurobiol.*, *55*(3), 257–332.

McKernan, M. G., & Shinnick-Gallagher, P. (1997). Fear conditioning induces a lasting potentiation of synaptic currents in vitro. *Nature*, *390*(6660), 607–611.

Miller, N. E. (1948). Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. Exp. Psychol.*, *38*, 89–101.

Miller, N. E. (1951). Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 435–472). New York: Wiley.

Miserendino, M. J., Sananes, C. B., Melia, K. R., & Davis, M. (1990). Blocking of acquisition but not expression of conditioned fear-potentiated startle by nmda antagonists in the amygdala. *Nature*, *345*(6277), 716–718.

Mogenson, G. J., Jones, D. L., & Yim, C. Y. (1980). From motivation to action: Functional interface between the limbic system and the motor system. *Prog. Neurobiol.*, *14*(2–3), 69–97.

Mowrer, O. H. (1947). On the dual nature of learning: A reinterpretation of "conditioning" and "problem solving." *Harvard Educ. Rev.*, *17*, 102–148.

Mowrer, O. H., & Lamoreaux, R. R. (1946). Fear as an intervening variable in avoidance conditioning. *J. Comp. Psychol.*, *39*, 29–50.

Nicola, S. M. (2007). The nucleus accumbens as part of a basal ganglia action selection circuit. *Psychopharmacology (Berl.)*, *191*(3), 521–550.

Niv, Y., Joel, D., & Dayan, P. (2006). A normative perspective on motivation. *Trends Cogn. Sci.*, *10*(8), 375–381.

Overmier, J. B., & Lawry, J. A. (1979). Pavlovian conditioning and the mediation of avoidance behavior. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 13, pp. 1–55). New York: Academic Press.

Pape, H. C., & Stork, O. (2003). Genes and mechanisms in the amygdala involved in the formation of fear memory. *Ann. NY Acad. Sci.*, *985*, 92–105.

Pare, D., Quirk, G. J., & LeDoux, J. E. (2004). New vistas on amygdala networks in conditioned fear. *J. Neurophysiol.*, *92*(1), 1–9.

Parkinson, J. A., Olmstead, M. C., Burns, L. H., Robbins, T. W., & Everitt, B. J. (1999). Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive Pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by d-amphetamine. *J. Neurosci.*, *19*(6), 2401–2411.

Parkinson, J. A., Robbins, T. W., & Everitt, B. J. (2000). Dissociable roles of the central and basolateral amygdala in appetitive emotional learning. *Eur. J. Neurosci.*, *12*(1), 405–413.

Pascoe, J. P., & Kapp, B. S. (1985). Electrophysiological characteristics of amygdaloid central nucleus neurons during Pavlovian fear conditioning in the rabbit. *Behav. Brain Res.*, *16*(2–3), 117–133.

Pavlov, I. P. (1927). *Conditioned reflexes*. New York: Dover.

Petrovich, G. D., & Swanson, L. W. (1997). Projections from the lateral part of the central amygdalar nucleus to the postulated fear conditioning circuit. *Brain Res.*, *763*(2), 247–254.

Pitkänen, A. (2000). Connectivity of the rat amygdaloid complex. In J. P. Aggleton (Ed.), *The amygdala: A functional analysis* (pp. 31–115). Oxford, UK: Oxford University Press.

Pitkanen, A., Pikkarainen, M., Nurminen, N., & Ylinen, A. (2000). Reciprocal connections between the amygdala and the hippocampal formation, perirhinal cortex, and postrhinal cortex in rat: A review. *Ann. NY Acad. Sci.*, *911*, 369–391.

Pitkänen, A., Savander, V., & LeDoux, J. E. (1997). Organization of intra-amygdaloid circuitries in the rat: An emerging framework for understanding functions of the amygdala. *Trends Neurosci.*, *20*(11), 517–523.

Quirk, G. J., Armony, J. L., & LeDoux, J. E. (1997). Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron*, *19*(3), 613–624.

Quirk, G. J., Garcia, R., & Gonzalez-Lima, F. (2006). Prefrontal mechanisms in extinction of conditioned fear. *Biol. Psychiatry*, *60*(4), 337–343.

Quirk, G. J., Repa, C., & LeDoux, J. E. (1995). Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: Parallel recordings in the freely behaving rat. *Neuron*, *15*(5), 1029–1039.

Repa, J. C., Muller, J., Apergis, J., Desrochers, T. M., Zhou, Y., & LeDoux, J. E. (2001). Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nat. Neurosci.*, *4*(7), 724–731.

Rescorla, R. A., & Solomon, R. L. (1967). Two process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychol. Rev.*, *74*, 151–182.

Robbins, T. W., Giardini, V., Jones, G. H., Reading, P., & Sahakian, B. J. (1990). Effects of dopamine depletion from the caudate-putamen and nucleus accumbens septi on the acquisition and performance of a conditional discrimination task. *Behav. Brain Res.*, *38*, 243–261.

Roberts, A. C., & Glanzman, D. L. (2003). Learning in aplysia: Looking at synaptic plasticity from both sides. *Trends Neurosci.*, *26*(12), 662–670.

Rodrigues, S. M., Schafe, G. E., & LeDoux, J. E. (2001). Intra-amygdala blockade of the nr2b subunit of the nmda receptor disrupts the acquisition but not the expression of fear conditioning. *J. Neurosci.*, *21*(17), 6889–6896.

Rodrigues, S. M., Schafe, G. E., & LeDoux, J. E. (2004). Molecular mechanisms underlying emotional learning and memory in the lateral amygdala. *Neuron*, *44*(1), 75–91.

Rogan, M. T., & LeDoux, J. E. (1995). Ltp is accompanied by commensurate enhancement of auditory-evoked responses in a fear conditioning circuit. *Neuron*, *15*(1), 127–136.

Rogan, M. T., Staubli, U. V., & LeDoux, J. E. (1997). Fear conditioning induces associative long-term potentiation in the amygdala. *Nature*, *390*(6660), 604–607.

Rolls, E. T. (2004). The functions of the orbitofrontal cortex. *Brain Cogn.*, *55*(1), 11–29.

Romanski, L. M., LeDoux, J. E., Clugnet, M. C., & Bordi, F. (1993). Somatosensory and auditory convergence in the lateral nucleus of the amygdala. *Behav. Neurosci.*, *107*(3), 444–450.

Rorick-Kehn, L. M., & Steinmetz, J. E. (2005). Amygdalar unit activity during three learning tasks: Eyeblink classical conditioning, Pavlovian fear conditioning, and signaled avoidance conditioning. *Behav. Neurosci.*, *119*(5), 1254–1276.

Rosen, J. B. (2004). The neurobiology of conditioned and unconditioned fear: A neurobehavioral system analysis of the amygdala. *Behav. Cogn. Neurosci. Rev.*, *3*(1), 23–41.

Rosenkranz, J. A., & Grace, A. A. (2002). Dopamine-mediated modulation of odour-evoked amygdala potentials during pavlovian conditioning. *Nature, 417*, 282–287.

Royer, S., Martina, M., & Pare, D. (1999). An inhibitory interface gates impulse traffic between the input and output stations of the amygdala. *J. Neurosci.*, *19*(23), 10575–10583.

Rumpel, S., LeDoux, J., Zador, A., & Malinow, R. (2005). Postsynaptic receptor trafficking underlying a form of associative learning. *Science*, *308*(5718), 83–88.

Sah, P., Faber, E. S., Lopez De Armentia, M., & Power, J. (2003). The amygdaloid complex: Anatomy and physiology. *Physiol. Rev.*, *83*(3), 803–834.

Salamone, J. D., Correa, M., Farrar, A., & Mingote, S. M. (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology (Berl.)*, *191*(3), 461–482.

Salamone, J. D., Correa, M., Mingote, S., & Weber, S. M. (2003). Nucleus accumbens dopamine and the regulation of effort in food-seeking behavior: Implications for studies of natural motivation, psychiatry, and drug abuse. *J. Pharmacol. Exp. Ther.*, *305*(1), 1–8.

Samson, R. D., Duvarci, S., & Pare, D. (2005). Synaptic plasticity in the central nucleus of the amygdala. *Rev. Neurosci.*, *16*(4), 287–302.

Samson, R. D., & Pare, D. (2005). Activity-dependent synaptic plasticity in the central nucleus of the amygdala. *J. Neurosci.*, *25*(7), 1847–1855.

Sarter, M. F., & Markowitsch, H. J. (1985). Involvement of the amygdala in learning and memory: A critical review, with emphasis on anatomical relations. *Behav. Neurosci.*, *99*, 342–380.

Schiller, D., & Weiner, I. (2004). Lesions to the basolateral amygdala and the orbitofrontal cortex but not to the medial prefrontal cortex produce an abnormally persistent latent inhibition in rats. *Neuroscience*, *128*(1), 15–25.

Schneiderman, N., Francis, J., Sampson, L. D., & Schwaber, J. S. (1974). CNS integration of learned cardiovascular behavior. In L. V. DiCara (Ed.), *Limbic and autonomic nervous system research* (pp. 277–309). New York: Plenum.

Schoenbaum, G., Gottfried, J. A., Murray, E. A., & Ramus, S. J. (2007). Linking affect to action: Critical contributions

of the orbitofrontal cortex. Preface. *Ann. NY Acad. Sci.*, *1121*, xi–xiii.

Selden, N. R., Everitt, B. J., Jarrard, L. E., & Robbins, T. W. (1991). Complementary roles for the amygdala and hippocampus in aversive conditioning to explicit and contextual cues. *Neuroscience*, *42*(2), 335–350.

Seligman, M. E., & Johnston, J. C. (1973). A cognitive theory of avoidance learning. In F. J. McGuigan & D. B. Lumsden (Eds.), *Contemporary approaches to conditioning and learning* (pp. 69–110). Oxfored, UK: V.H. Winston & Sons.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.

Solomon, R. L., & Wynne, L. C. (1954). Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychol. Rev.*, *61*, 353–385.

Sotres-Bayon, F., Cain, C. K., & LeDoux, J. E. (2006). Brain mechanisms of fear extinction: Historical perspectives on the contribution of prefrontal cortex. *Biol. Psychiatry*, *60*(4), 329–336.

Swanson, L. W., & Petrovich, G. D. (1998). What is the amygdala? *Trends Neurosci.*, *21*(8), 323–331.

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human Pavlovian-instrumental transfer. *J. Neurosci.*, *28*(2), 360–368.

Thompson, R. F. (1976). The search for the engram. *Am. Psychol.*, *31*, 209–227.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychol. Monogr.*, *2*, 109.

Tinbergen, N. (1951). *The study of instinct*. New York: Oxford University Press.

Wadenberg, M. L., Ericson, E., Magnusson, O., & Ahlenius, S. (1990). Suppression of conditioned avoidance behavior by the local application of (–)sulpiride into the ventral, but not the dorsal, striatum of the rat. *Biol. Psychiatry*, *28*(4), 297–307.

Walker, D. L., & Davis, M. (2002). The role of amygdala glutamate receptors in fear learning, fear-potentiated startle, and extinction. *Pharmacol. Biochem. Behav.*, *71*(3), 379–392.

Walters, E. T., Carew, T. J., & Kandel, E. R. (1979). Classical conditioning in Aplysia californica. *Proc. Natl. Acad. Sci. USA*, *76*(12), 6675–6679.

Wang, H., Hu, Y., & Tsien, J. Z. (2006). Molecular and systems mechanisms of memory consolidation and storage. *Prog. Neurobiol.*, *79*(3), 123–135.

Watson, J. B. (1929). *Behaviorism*. New York: W. W. Norton.

Weiskrantz, L. (1956). Behavioral changes associated with ablation of the amygdaloid complex in monkeys. *J. Comp. Physiol. Psychol.*, *49*, 381–391.

Weisskopf, M. G., Bauer, E. P., & LeDoux, J. E. (1999). L-type voltage-gated calcium channels mediate NMDA-independent associative long-term potentiation at thalamic input synapses to the amygdala. *J. Neurosci.*, *19*(23), 10512–10519.

Wilensky, A. E., Schafe, G. E., Kristensen, M. P., & LeDoux, J. E. (2006). Rethinking the fear circuit: The central nucleus of the amygdala is required for the acquisition, consolidation, and expression of Pavlovian fear conditioning. *J. Neurosci.*, *26*(48), 12387–12396.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.*, *7*(6), 464–476.

Yin, J. C., & Tully, T. (1996). CREB and the formation of long-term memory. *Curr. Opin. Neurobiol.*, *6*(2), 264–268.

Zimmerman, J. M., Rabinak, C. A., McLachlan, I. G., & Maren, S. (2007). The central nucleus of the amygdala is essential for acquiring and expressing conditional fear after overtraining. *Learn. Memory*, *14*(9), 634–644.

62 Interactions of Emotion and Attention in Perception

PATRIK VUILLEUMIER AND TOBIAS BROSCH

ABSTRACT Interactions between brain systems involved in emotion and attention have attracted a great deal of interest, as both may contribute to regulating behavior and awareness by enhancing the representation of sensory information relevant to the individual. This chapter reviews recent research derived from cognitive sciences and brain imaging that reveals a modulation of early perceptual pathways by emotional signals and suggests a crucial role for the amygdala in imposing direct feedback influences on sensory cortical areas that can boost perception and attention for emotionally relevant stimuli. While such emotional effects may arise in parallel to top-down influences mediated by attentional systems, a different task or context is likely to modulate the efficacy of such boosting, but such modulations still remain poorly known and debated. In addition, although emotional influences from the amygdala on perceptual processing have most often been studied with fear-related stimuli, similar effects might also arise for positive or arousing stimuli that are self-relevant.

To successfully move about in the world and respond to challenges, we have to rapidly detect unexpected changes and important information within our environment. But because of the brain's capacity limits, only a subset of all incoming stimuli can be selected for more in-depth processing, affording subsequent access to other systems such as memory, motor control, and conscious awareness. This competition for neural processing resources is influenced by several factors related to the observer and/or the stimulus. Top-down factors such as expectations and current goals can bias the competition and boost the neural representation of specific stimuli, for example, in searching for a given object in a scene. This voluntary control, termed *endogenous attention*, involves a modulation of sensory pathways by frontoparietal regions (Driver & Frackowiak, 2001). Conversely, physically intense or deviant properties of stimuli, such as their brightness or loudness, may elicit enhanced responses in the sensory stream and trigger reflexive orienting to these stimuli, corresponding to *exogenous attention*.

PATRIK VUILLEUMIER Department of Neuroscience, University Medical Center, Geneva; Swiss National Center for Affective Sciences, University of Geneva, Geneva, Switzerland
TOBIAS BROSCH Department of Psychology, University of Geneva; Swiss National Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Remarkably, the emotional relevance of a stimulus appears to play a similar and complementary role for controlling the allocation of processing resources for perception and awareness (Vuilleumier, 2005). A central function of emotions is to determine the relevance of a stimulus for well-being and survival and then coordinate an appropriate behavioral response (Scherer, 2001). Efficient processing of emotionally relevant stimuli is obviously highly adaptive, as unexpected events with emotional significance should be noticed more readily and, once detected, become the focus of attention, evaluation, and action. Compelling evidence indicates that this form of *emotional attention* is driven by specific neural mechanisms, including the amygdala. Hence both attention and emotion may contribute to regulate perception and access to conscious awareness, though via partly distinct neural mechanisms. Thus, emotion processing not only serves to imbue our experiences with affective flavors and feelings, but also directly shapes the content of awareness itself. In this chapter, we will provide a general overview of the reciprocal interplay between attention and emotion, considering both brain imaging and behavioral evidence, and will describe our current knowledge of underlying neural circuits.

Enhanced neural processing of emotional information

From everyday experience, we know that emotional stimuli have a special role in perception. Smiling people, attractive faces, one's own name or the name of a loved one overheard at a party, pictures of bloody mutilations, and screams all appear more salient and attention grabbing than neutral stimuli do. In line with these subjective impressions, brain imaging studies in humans have demonstrated increased neural responses to a great variety of emotional stimuli relative to comparable neutral stimuli, using different experimental paradigms and techniques such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET), electroencephalography (EEG), or magnetoencephalography (MEG).

In the visual domain, an emotional "boosting" of neural responses to emotional compared to neutral stimuli has been observed in early visual areas, including primary visual

cortex, as well as in higher-level regions associated with object and face recognition. Pictures of scenes with emotional content produce greater activation in the lateral occipital cortex, faces with emotional expressions produce increases in the fusiform face area (FFA), and emotional body expressions activate both the fusiform and extrastriate body areas (FBA and EBA). Likewise, vocal and nonvocal sounds with emotional significance may evoke higher neural responses in auditory cortical areas than similar but more mundane sounds will. These findings suggest a selective modulation by emotion of brain regions involved in processing the specific stimulus category (see figure 62.1). Increased responses in visual or auditory cortex are also obtained for previously neutral stimuli after aversive conditioning. This suggests that such boosting is not (exclusively) due to intrinsic sensory features of the stimuli but directly relate to their emotional meaning.

Similarly, electrophysiological measures with EEG or MEG have shown enhanced cortical responses to emotional scenes, facial expressions, or fear-conditioned stimuli (for reviews, see Olofsson, Nordin, Sequeira, & Polich, 2008; Vuilleumier & Pourtois, 2007). Such effects may arise at several stages of perceptual processing, including early visual components (100–200 ms after stimulus onset) generated in striate and extrastriate cortex (such as C1, P1, or N1) as well as subsequent components such as the N170 or posterior negativities associated with object recognition processes in higher-level regions. Emotional increases also arise at longer latencies poststimulus onset, such as modulations of the P3 component and sustained late positive potentials (LPP), which may reflect more elaborate affective and cognitive evaluations of the stimuli, autonomic arousal, and/or memory formation. Source localization analyses suggest that these late ERP components reflect the activity of a widespread cortical network including prefrontal, cingulate, and parietal regions. LPP amplitude in EEG is also correlated with the magnitude of fMRI response in occipitotemporal cortex.

Several PET and fMRI studies have also reported that these cortical increases were significantly correlated with amygdala responses; that is, the more the amygdala was sensitive to the emotional meaning in a visual stimulus, the greater the responses of visual areas to this stimulus. Likewise, in conditioning studies, enhanced activation toward conditioned stimuli is usually accompanied by concomitant activation in the amygdala. This has led to the idea that the amygdala may play a critical role in driving such cortical increases, consistent with physiological and evolutionary considerations (Öhman, 2005) suggesting a key function of the amygdala for the rapid detection and response to potential dangers (see below). Furthermore, a vast majority of imaging studies demonstrating emotional influences on cortical perceptual responses used threat-related cues such as fearful or angry faces as well as aversively conditioned stimuli, consistent with the central role of the amygdala in fear processing and fear learning (LeDoux, 1996). However, a similar enhancement of cortical responses has been observed in human visual and auditory cortex for positive visual scenes, erotica, pictures of babies, or joyful voices. These positive stimuli also increase amygdala activation, suggesting that the emotional processes underlying these effects might not be exclusively sensitive to fear but could more generally be sensitive to emotionally relevant or arousing information (Sander, Grafman, & Zalla, 2003).

Taken together, imaging and electrophysiology data show that emotion can boost perceptual processing at both early sensory and higher-level cortical stages, presumably affording a more robust representation and preferential access of these stimuli to further cognitive processing and awareness.

Behavioral effects of emotion on perceptual processing

In keeping with neuroimaging and electrophysiological evidence of enhanced neural responses, behavioral findings indicate that perception is facilitated and attention is prioritized for emotional information. Emotional stimuli may draw attention more quickly and impede attentional disengagement longer than neutral stimuli, which, depending on the task, can improve behavioral performance when targets are emotionally relevant but also lead to interference when an emotional stimulus competes with a nonemotional target.

Examples for behavioral facilitation by emotion come from visual search studies, in which the detection of a target among distractors is typically faster when the target is emotional rather than neutral (e.g., Öhman, Flykt, & Esteves, 2001). Although this has more often been shown with negative or threat-related stimuli such as angry faces, snakes, or guns, similar effects have sometimes been reported with positive or appetitive stimuli. Conversely, emotional distractors may impair search for a nonemotional target. More efficient searches do not imply that emotional stimuli are processed without attention or "pop out" like targets defined by salient feature differences (e.g., color), as was often interpreted in earlier studies. Instead, attention appears to be preferentially guided toward emotional information, reflecting biases in the allocation of attention rather than a shortcut to conscious perception. Thus, detection time slopes are typically shallower when targets are emotional rather than neutral but do not remain flat irrespective of the number of items in the search display.

Although these data suggest that emotion may drive attention, one cannot exclude the concept that the effects are due to associated stimulus or task characteristics and not direct

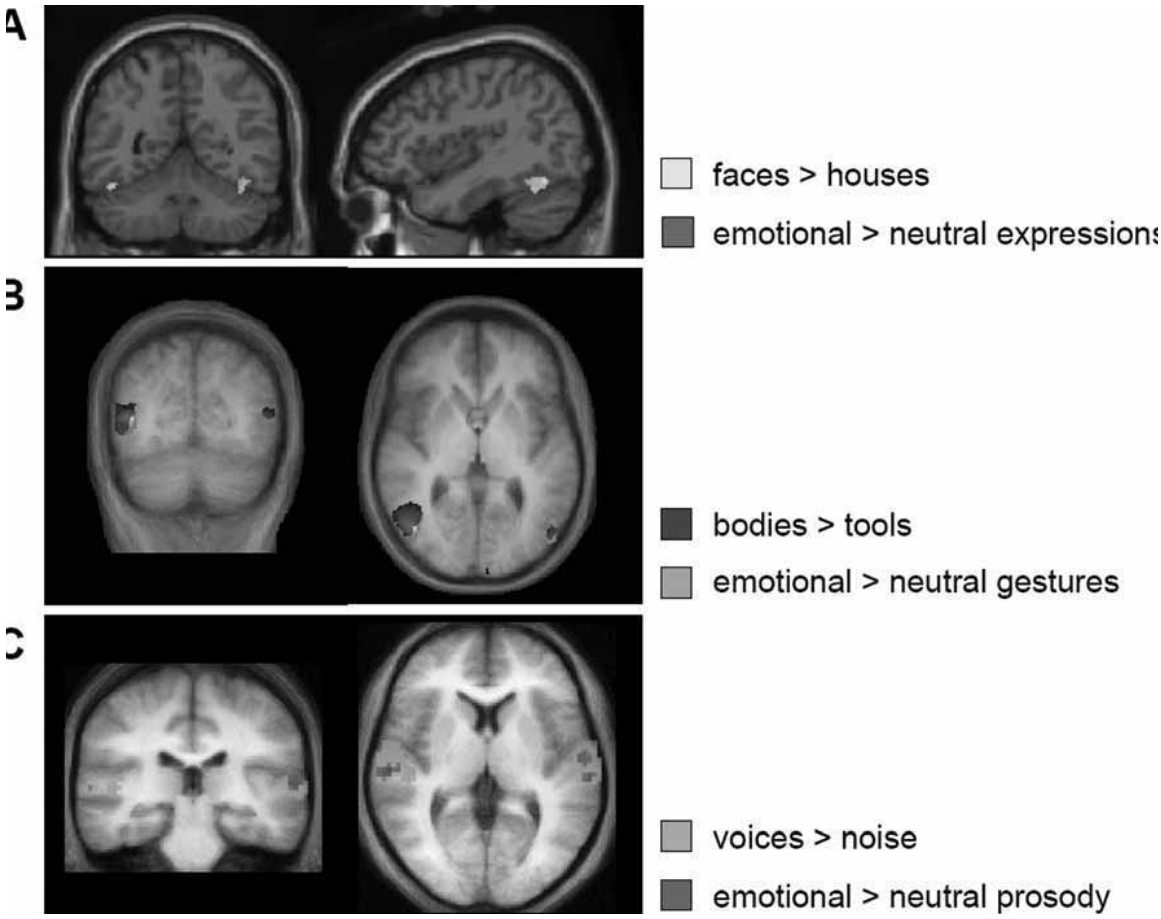


FIGURE 62.1 Emotional enhancement of neural responses in fMRI. (*A*) Faces with a fearful relative to neutral expression produce increased activation in fusiform cortex, overlapping with the fusiform area selectively activated by faces as compared with houses. (From Vuilleumier et al., 2001.) (*B*) Bodies with dynamic gestures expressing various emotions (fear, anger, happiness, or disgust) produce increased activation in lateral occipital area cortex, overlapping with the extrastriate area selectively activated by bodies as compared with tools. (From Peelen, Atkinson, Andersson, & Vuilleumier, 2007.) (*C*) Voices with angry prosody produce increased activation in temporal cortex, overlapping with an area in superior temporal gryus selectively activated by human voices as compared with noises with similar acoustic energy. (From Grandjean et al., 2005.) (See color plate 78.)

effects of emotional appraisal. For example, the degree of attentional capture by an emotional stimulus does not always correspond to the strength of affective evaluations for the same stimulus when measured by implicit tests such as affective priming (Purkis & Lipp, 2007). However, in other tasks, ratings of emotional intensity correlate with degree of response facilitation (Brosch, Sander, & Scherer, 2007). A role of emotional processes is also supported by the findings that attentional biases in search (and other) tasks can be modulated by individual state and trait differences related to emotion. For example, attentional bias toward relevant threatening information is often enhanced in people with specific phobia: Attention is directed faster to pictures of snakes than to pictures of spiders in snake phobics but vice versa in spider phobics (Öhman et al., 2001). Furthermore, depressive individuals show an increased bias toward negative information, whereas high levels of optimism are associated with stronger attentional bias toward positive information. State differences related to current mood may also influence attentional biases toward positive or negative information (Smith et al., 2006). These individual differences strongly suggest that prioritized attention is determined by an appraisal of the emotional meaning and personal relevance of a stimulus rather than just salient sensory features. However, it is still possible that such effects reflect a greater sensitivity or tuning of perceptual systems to some critical features, perhaps subsequent to emotional experience and learning (Weinberger, 1998).

Enhanced performance induced by emotional meaning has been observed in other visual tasks, such as the attentional blink. In the latter, the detection of a target word in a rapid serial visual stream (items appearing successively at fixation at about 10 Hz) is impaired when it occurs shortly after another target. However, this deficit is greatly attenuated for emotional stimuli (e.g., Anderson & Phelps, 2001). Conversely, it may increase for a second neutral target

following an emotional one, suggesting that the emotional meaning of items tend to either grab or divert attention in situations in which resources cannot be equally deployed for every successive stimulus.

Examples of disruption of performance by emotion come from variants of the Stroop interference task, in which participants name the color of a word that can be emotional or neutral. Slower response times to emotional items are generally interpreted to reflect enhanced attentional capture by their emotional meaning. Although a negativity bias has commonly been reported, in the form of more interference by negative than by positive or neutral words, particularly in clinical populations with anxiety disorders, there is also a larger interference for words that are related to individually relevant topics. This has been demonstrated, for example, for spider phobics, social phobics, rape victims, and posttraumatic stress disorder (PTSD) patients (J. M. G. Williams, Mathews, & MacLeod, 1996).

Once attention has been drawn to and engaged by emotional stimuli, it may also dwell longer at their location and facilitate the processing of subsequent stimuli. Such orienting effects have been demonstrated by using the dot probe task, in which participants have to respond to a target (e.g., a dot) that replaces one of two simultaneously presented cues, one cue being emotionally significant (e.g., a fearful or angry face) and the other being neutral. Typical results show faster responses to targets that replace the emotional cue (valid trial) rather than the neutral cue (invalid trial). These effects may reflect either a facilitation of orienting attention toward the emotional stimulus or a greater difficulty in disengaging attention from it (Koster, Crombez, Verschuere, & De Houwer, 2004). In addition to faster response times, emotional cueing may also increase contrast sensitivity for the subsequent target (Phelps, Ling, & Carrasco, 2006). These effects are essentially exogenous and reflexive, since they occur despite the fact that the cue is not predictive of the target location. They may arise with both negative and positive emotional cues (Brosch, Sander, Pourtois, & Scherer, 2008) and operate even across sensory modalities (i.e., for visual targets following auditory cues) (Brosch, Grandjean, Sander, & Scherer, 2008), suggesting that the prioritization of emotionally relevant stimuli is organized supramodally across multiple sensory channels.

Thus, both behavioral and imaging data converge to indicate that perceptual processing is enhanced for emotionally relevant information. Such enhancement may affect performance under conditions of limited resources, in which stimuli in a cluttered scene or in rapid succession must compete for attention, allowing better detection or faster responses to emotional stimuli. Whereas attentional control is generally thought to depend on top-down signals from frontal and parietal areas boosting the neural processing of task-relevant stimuli, the influence of emotion appears to produce a similar boosting on perceptual pathways to increase the representation of affectively relevant stimuli and bias attention toward them, but the neural mechanisms may implicate partly distinct brain circuits, as will be described below.

Neural circuits underlying emotional attention

The amygdala is thought to be critically involved in processing the emotional relevance of stimuli but also in mediating the influence of emotion on perception. Its dense reciprocal connections with widespread regions in the cortex, including all stages along the perceptual pathways as well as prefrontal regions (Amaral, Behniea, & Kelly, 2003), enable it to receive rich sensory information and to exert both direct and indirect feedback on sensory pathways to boost the processing of emotional stimuli (see figure 62.2). Anatomical studies in the macaque show that direct feedback projections from the amygdala on visual cortex are topographically organized, with denser projections to rostral, high-level areas than to caudal, low-level areas. Most of these projections arise from the basal nucleus of the amygdala but with a progressive gradient, such that its dorsal part (magnocellular, Bmc) predominantly targets earlier areas (such as V1) and its more

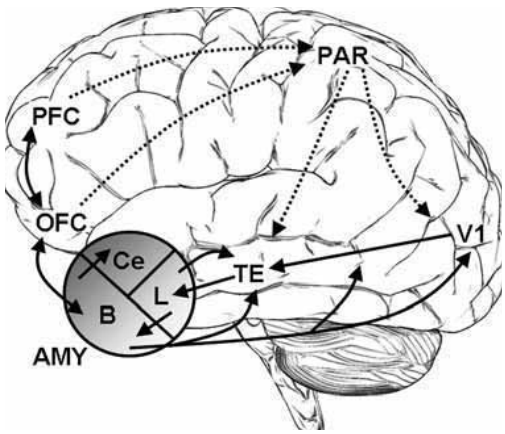


Figure 62.2 Reciprocal pathways between emotional and attentional control. Feedback from lateral (L) and basal (B) nuclei from amygdala (AMY) can amplify neural representations of emotionally relevant information at different stages along sensory cortical areas (as illustrated here for the visual system, with early primary visual cortex, V1, and later visual areas in inferotemporal cortex, TE.). While the lateral nucleus receives and projects mainly to higher-level areas, the basal nucleus projects to all stages of sensory processing. Top-down signals from parietal cortex (PAR) on sensory areas may focus attentional resources on the location of emotional events. Amygdala feedback loops may be modulated by influences from orbitofrontal cortex (OFC), as well as by interconnections with prefrontal areas (PFC). Amygdala output via the central nucleus (Ce) can also activate other brain systems, including cholinergic projections from forebrain to parietal, as well as frontal and sensory cortical regions. These projections are not shown here.

ventral part (intermediate, Bi) predominantly reaches later areas (such as TEO and TE). By contrast, most visual inputs to the amygdala project to the lateral nucleus, which then projects to the basal nucleus and back to higher-level areas only (such as TE). A similar topographical organization probably exists in the auditory domain.

Direct evidence for a modulatory influence of this amygdala feedback loop on cortical responses in humans is provided by fMRI results showing that amygdala lesions may impair functional activation of intact visual areas to emotional faces. In a visual discrimination task comparing responses to fearful and neutral faces (presented at either attended or ignored locations), healthy controls typically show increased activation in fusiform cortex to fearful faces relative to neutral faces, irrespective of whether attention is focused on the faces or concurrent house pictures instead (Vuilleumier, Armony, Driver, & Dolan, 2001). However, this increase is not observed in patients with amygdala damage due to medial temporal lobe sclerosis, in contrast with patients whose sclerosis is restricted to the hippocampus and does not affect the amygdala (Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004). These findings indicate that the amygdala plays a causal role in the enhanced processing of emotional stimuli within distant cortical sites. Preliminary ERP results have also pointed to functional anomalies in visual responses to emotional faces in patients with amygdala lesions, affecting both the early P1 component and later components that arise at 500 ms after stimulus onset.

Furthermore, an impact of amygdala lesion on visual performance has also been demonstrated behaviorally in patients with temporal lobe resection (Anderson & Phelps, 2001). Those with left or bilateral damage to the amygdala did not show the normal facilitation of detection for emotional relative to neutral words in an attentional blink paradigm, even though the patients still understood normally the affective meaning. This deficit provides further evidence that affective influences on attentional processes depend on amygdala function.

Amygdala outputs also project to cholinergic basal forebrain nuclei, which in turn can modulate perception through their widespread connections with many cortical areas. However, imaging results have shown that cholinergic stimulation may modulate orbitofrontal and parietal regions, without significant changes in the normal emotional boosting of sensory (e.g., visual) areas (Bentley, Vuilleumier, Thiel, Driver, & Dolan, 2003).

In addition, emotional stimuli can produce indirect effects by modulating activity within the frontoparietal attention network. This was observed, for example, in the reflexive spatial orienting induced by peripheral emotional stimuli in a dot probe task (Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2006), in which a neutral visual target (e.g., a dot) is presented following a pair of faces (e.g., one fearful and one neutral), replacing either the emotional or the neutral face (valid or invalid trials, respectively; see above). Greater fMRI activation is observed in the intraparietal sulcus (IPS) when targets are preceded by a fearful face than a neutral face, consistent with enhanced attentional orienting and faster detection of targets on valid trials. This increased activation of IPS on valid trials contrasted with strongly reduced activation on invalid trials in which targets are presented in the ipsilateral visual field after an emotional face on the contralateral side, suggesting that IPS may become unresponsive to stimuli in the ipsilateral hemifield subsequent to the enhanced focusing of attention on the contralateral side. These neural effects corroborate behavioral findings that emotional stimuli may not only draw but also hold attention to their location (Fox, Russo, & Dutton, 2002).

In keeping with these fMRI results, ERP recordings during the dot probe paradigm (Brosch, Sander, et al., 2008; Pourtois, Grandjean, Sander, & Vuilleumier, 2004) show higher amplitude of the P1 visual potential to targets that replace an emotional rather than a neutral face. As the P1 is generated in extrastriate occipital cortex and modulated by spatial attention, these data further demonstrate that emotional cues may bias spatial attention and enhance visual processing for subsequent stimuli at the same location. Detailed spatiotemporal analysis of the ERPs suggests that this enhancement of P1 is preceded by a modulation of parietal activity that correlates with the magnitude of P1 increases (Pourtois, Thut, Grave de Peralta, Michel, & Vuilleumier, 2005), suggesting that this earlier parietal activity may be responsible for generating top-down influences on visual cortex following the presentation of the emotional cue.

Taken together, these data show that emotional signals may not only activate the amygdala and related emotional brain systems, but also directly affect sensory cortices to modulate the neural representation of a stimulus and additionally influence frontoparietal mechanisms responsible for orienting and shifting attention in space. Thus, subsequent information arising at the same location as emotional cues will also benefit from enhanced processing resources.

The automaticity of emotional attention

Most of the emotional influences on behavioral and neural processes described above arise in conditions in which emotional meaning itself is not directly relevant to the task and therefore appears to be processed automatically without intention or even without awareness. For instance, the emotional Stroop interference demonstrates that word meaning is extracted involuntarily, although this is not required and in fact is counterproductive with respect to the task goal. Likewise, in the dot probe task or visual search (see above),

attention seems reflexively oriented toward emotional cues even when these are not predictive of target location.

However, there are still debates about the exact degree of "automaticity" of emotional processing and emotional attention (Pessoa, 2005; Vuilleumier & Driver, 2007). Automaticity implies several different functional features (related to intentionality, controllability, rapidity, and awareness) that do not necessarily co-occur (Moors & De Houwer, 2006), but much remains to be done to disentangle these different aspects in relation to emotional processing. Three main questions that are often subsumed under this issue concern (1) whether emotional responses in the amygdala require cortical processing or can be triggered by a subcortical shortcut, (2) whether emotional processing has to compete for general attentional resources or takes place irrespective of task demands, and (3) whether attentional effects induced by emotion are totally independent from attentional processes mediated by frontoparietal networks.

The first question concerning the route of emotional inputs to the amygdala derives from fear-conditioning experiments in rats (LeDoux, 1996), showing that auditory information may reach the amygdala via a short latency pathway from the thalamus, the so-called low road, even after interruption of cortical pathways. By analogy, a visual shortcut (via collicular-pulvinar pathways) has been proposed in humans (Morris, Öhman, & Dolan, 1999) to account for the findings that healthy subjects may show amygdala activation to masked fearful faces that are not perceived consciously and that patients who are blind after destruction of visual cortex may still discriminate emotional stimuli and show increased amygdala activation to emotional facial expressions (Pegna, Khateb, Lazeyras, & Seghier, 2005). Subcortical inputs might also reach the amygdala via direct pulvinar projection to extrastriate areas. Although there is no direct evidence in humans that a subcortical route would be truly faster than the cortical route, collicular-pulvinar pathways are known to carry only magnocellular visual inputs (not parvocellular), which convey fast but coarse, low-spatial frequency information throughout the visual system (Vuilleumier, Armony, Driver, & Dolan, 2003).

Alternatively to this two-pathway mechanism, a two-stage mechanism might also allow rapid inputs to reach the amygdala through a first feedforward sweep of cortical activation (Bullier, 2001). Amygdala responses could be triggered by a limited amount of information, such as low spatial frequency cues conveyed by magnocellular visual inputs to the cortex, or simple "diagnostic" features of emotional stimuli (e.g., wide-open eyes in fearful faces). This initial emotional appraisal could then provide subsequent feedback signals to cortical areas, where further processing may take place over longer periods on the basis of more complex sensory information (e.g., high-spatial frequency, parvocellular inputs). Both the two-pathway or two-stage mechanisms would account for ERP data demonstrating that the emotional value of some stimuli (e.g., facial expressions) can be encoded after short latencies (e.g., 100–140 ms), prior to higher-level processes associated with perceptual or categorical encoding (e.g., 170–200 ms). Recent research using MEG and dynamic causal modeling (Rudrauf et al., 2008) has provided support to the two-pathway hypothesis, by showing that a network including a fast subcortical pathway could better explain patterns of activation to emotional information. However, more direct neurophysiological evidence is still needed to disentangle the differences between these two hypotheses.

A second question on automaticity is whether limitation in attentional resources or awareness may reduce processing of emotional stimuli, like other classes of stimuli, or whether emotional stimuli are privileged or immune to general attention control. On the one hand, several imaging studies have demonstrated parallel influences of emotion and attention on brain responses. By manipulating attention and emotion separately in fMRI studies, it was found that the response to faces in visual cortex (Vuilleumier et al., 2001) or to voices in auditory cortex (Grandjean et al., 2005) could be modulated by each factor independently. In the former study, participants were presented with pairs of faces and pairs of houses and had to focus attention on one pair only (either faces or houses) while ignoring the other (to make a difficult same/different identity judgment). Faces were either both fearful or both neutral. As was expected, face-sensitive regions in fusiform cortex showed increased activation when attention was directed to pairs of faces rather than to houses; more important, fusiform activity was also greater for fearful faces than for neutral faces, both when faces were task-relevant and when they were ignored. In other words, fear expression could boost fusiform activity in a parallel and additive manner to the modulation by spatial attention on the same region. Similar results have been obtained by using an auditory dichotic task, in which voice-selective regions in superior temporal sulcus were not only modulated by voluntary attention (with greater responses when focusing on the contralateral than ipsilateral ear), but also modulated by emotional prosody (with greater responses to angry than neutral voices), irrespective of the side of the angry voices (i.e., in the attended or unattended ear). Furthermore, in the face-house task, amygdala lesions can abolish the modulation by emotion while preserving modulation by attention (Vuilleumier et al., 2004). These data demonstrate that emotional enhancement of face or voice processing may arise over and above a concomitant influence of endogenously driven attention and that such emotional response may still persist when attention is diverted and cortical processing is reduced. Moreover, in both studies, amygdala activation was unaffected by spatial attention, consistent with its presumed role in driving emotional enhancement. A similar additive

pattern has been found with ERPs to complex emotional scenes presented at attended or unattended locations in visual hemifields (Keil, Moratti, Sabatinelli, Bradley, & Lang, 2005).

Nevertheless, it is possible that attentional and emotional effects may interact in some other brain areas (e.g., STS or V1) and that further diversion of attentional resources is necessary to suppress emotional processes. For instance, directing attention away from emotional stimuli may reduce amygdala responses in tasks in which attentional load is particularly high (Pessoa, Padmala, & Morland, 2005) and thus possibly also reduce the impact of emotion on perceptual processing. Although many studies have reported amygdala responses to emotional stimuli without attention or even without awareness (e.g., Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003; Jiang & He, 2006; Whalen et al., 1998), several others have reported reduced responses in such conditions (e.g., Pessoa, Kastner, & Ungerleider, 2002; Pessoa et al., 2005). However, in some cases, inattention or unawareness may reduce the differential response to emotional stimuli, owing to concomitant increases to neutral or positive stimuli (Anderson et al., 2003; Silvert et al., 2007; M. A. Williams, McGlone, Abbott, & Mattingley, 2005) rather than just decreases to emotional stimuli. In addition, individual differences may also influence the pattern of responses. In a study using the same face-house task as above, individuals with low state anxiety actually showed reduced amygdala activation to unattended fearful faces, possibly driven by prefrontal attentional control mechanisms (Bishop, Duncan, Brett, & Lawrence, 2004). Greater amygdala activity in high-state anxiety might reflect weaker attentional control and/or amplification of feedback loops within the amygdala subnuclei under the influence of prefrontal regions (Kim et al., 2004).

The third issue concerning the independence of attentional and emotional influences on perception is supported not only by evidence of additive modulations in imaging studies (Keil et al., 2005; Vuilleumier et al., 2001) but also by neuropsychological observations indicating that emotional biases may still arise after brain lesions producing selective attentional deficits. Patients with right parietal or frontal damage may present with hemispatial neglect, characterized by failures in orienting attention to the contralesional/left space (Driver, Vuilleumier, & Husain, 2004); but the severity of neglect has been shown to be reduced for emotional stimuli relative to neutral stimuli, such as faces with angry or happy expressions, bodies with emotional gestures, pictures of spiders, or voices with various emotional prosodies (Grandjean, Sander, Lucas, Scherer, & Vuilleumier, 2008; Vuilleumier & Schwartz, 2001). Furthermore, a systematic analysis of brain lesions in these patients revealed that those with the largest gains in detection for emotional relative to neutral stimuli have lesions centered on lateral frontal and parietal regions, whereas those with weaker emotional biases have more frequent lesions in basal ganglia and orbitofrontal regions (for both emotional faces and emotional voices). Imaging results have also shown that fearful faces can still produce enhanced activation in intact visual cortex despite parietal damage and neglect and despite unawareness of such faces (Vuilleumier et al., 2002). These neuropsychological data add to the evidence that emotional influences on perception are not mediated by frontoparietal networks controlling spatial attention and further suggest that orbitofrontal regions might be implicated in mediating interactions between emotion and attention.

To sum up, although emotional stimuli may evoke reflexive and involuntary processing under many conditions, this does not preclude that such effects can be amplified or attenuated by cognitive or affective factors such as task load, anxiety, phobic traits, expectations, or prior experience. Involuntary monitoring of emotional stimuli and reflexive boosting of perceptual processes may reflect some "default" settings or intrinsic preparedness within neural pathways yet be adaptively shaped by various regulatory mechanisms that themselves can operate potentially with or without conscious control. However, the brain circuits that are responsible for such modulations remain to be determined, and their relationship to specific facets of automaticity still needs to be clarified.

Personality differences and cultural factors

The notion that interindividual differences influence attention and perception is not a new discovery. Bartlett (1932) noted that "temperament, interests and attitudes often direct the course and determine the content of perceiving." A number of behavioral results demonstrate that attentional biases to emotional information may depend on individual factors such as anxiety, optimism, positive or negative mood states, depression, or current goal states. Even cultural differences have been found to modulate attentional capture in the emotional Stroop task, with greater interference by the emotional semantics of words in American participants but greater interference by emotional prosody in Japanese participants (Ishii, Reyes, & Kitayama, 2003).

Although the neural substrates of these effects are unresolved, recent imaging studies have revealed different amygdala responses to emotional stimuli in relation to various personality traits, including anxiety, harm avoidance, extraversion, phobic fears, or attachment style (e.g., Bishop, Duncan, & Lawrence, 2004; Canli, Sivers, Whitfield, Gotlib, & Gabrieli, 2002; Sabatinelli, Bradley, Fitzsimmons, & Lang, 2005). A stronger "automatic" amygdala activation in anxious people may arise specifically under conditions of low cognitive or attentional load, together with reduced activation in areas associated with conflict monitoring and

executive control such as ACC and lateral PFC. Whereas state anxiety correlates with amygdala activation, trait anxiety correlates with relative deactivations in PFC and ACC, suggesting different effects of different aspects of anxiety (Bishop, Jenkins, & Lawrence, 2007). However, it remains to be determined whether the attentional biases that are described behaviorally in these conditions exert similar or different neural modulations, possibly implicating not only perceptual processing but also semantic processing, as well as memory retrieval, decision making, or thought.

Interactions between emotional attention and current behavioral goals

Rapid and reflexive processing of emotional stimuli without intention is adaptive to detect relevant information in the environment; however, it can also be distracting and detrimental to performance. For example, emotional distractors may impair working memory, accompanied by activation of amygdala and ventrolateral prefrontal cortex (VLPFC) but concurrent deactivation of task-related regions in dorsolateral prefrontal cortex (DLPFC) and lateral parietal cortex (Dolcos & McCarthy, 2006). A similar pattern has been observed during an attentional oddball task, with rare emotional and nonemotional pictures embedded in a stream of standard shapes. Amygdala and ventral frontal regions were activated by emotional stimuli, whether participants had to respond to them or not, while dorsal frontoparietal regions responded to target stimuli, irrespective of their emotionality. Anterior cingulate gyrus showed additive effects, responding to all emotional stimuli but responding more strongly when they were target (Fichtenholtz et al., 2004). Behavioral interference in the emotional Stroop task is also accompanied by activation of amygdala and rostral anterior cingulate cortex. ACC activation might reflect a processing conflict due to attention capture by emotional stimuli, which might then trigger activation in the lateral PFC to adjust and maintain top-down goal-related control.

Other studies have demonstrated a modulation of amygdala activity by behavioral goals, expectations, or emotion regulation efforts, which can perhaps determine the degree of emotional influences on sensory processing. However, amygdala activation toward negative information seems less amenable to attenuation by voluntary goals than does activation toward positive information (Cunningham, Van Bavel, & Johnsen, 2008). Different expectations can also produce different effects on amygdala responses to relevant stimuli. Single-cell recordings in monkeys suggest that some neuronal populations within the amygdala may be modulated by expectations of reward or punishment selectively, whereas others respond to any unexpected reinforcer independent of valence (Belova, Paton, Morrison, & Salzman, 2007). The effects of emotion regulation and expectations are presumably mediated by connections from orbitofrontal cortex (OFC), which may act as a gateway between emotional processes in the amygdala and representational memory systems in the prefrontal cortex (Roesch & Schoenbaum, 2006). OFC might thus strengthen or inhibit amygdala responses based on current context and internal state.

In summary, interactions between attention and emotion involve a large network that not only is centered on the amygdala, but also has many reciprocal connections with several prefrontal areas, including orbitofrontal and cingulate cortex as well as DLPFC (Cavada, Company, Tejedor, Cruz-Rizzolo, & Reinoso-Suarez, 2000). This network is well positioned to appraise the emotional relevance of stimuli by integrating representations of affective value with complex situational factors related to goals, expectations, experience, or personality and might serve to modulate both perceptual processes within sensory cortices and other cognitive processes within dorsal and lateral prefrontal areas. Through these interactions, by analogy with the notion of attentional sets (Folk, Remington, & Johnston, 1992), *emotional attentional sets* might sensitize neural pathways to stimuli that are emotionally relevant, given both their intrinsic meaning and the current state of the individual.

Conclusions

Recent research has provided us with a remarkable amount of new knowledge concerning the mechanisms by which perception and attention may be influenced by emotional processing. Such influences are implemented by a dynamic interplay between the amygdala and other brain regions, including sensory cortices as well as parietal and prefrontal areas. These emotional mechanisms complement *endogenous* and *exogenous* attentional systems that are known to select and organize sensory inputs based on voluntary behavioral goals and low-level physical salience and thus constitute a specialized neural system for *emotional* attention in the service of fast and adaptive response to highly relevant events.

REFERENCES

Amaral, D. G., Behniea, H., & Kelly, J. L. (2003). Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. *Neuroscience, 118*, 1099–1120.

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. (2003). Neural correlates of the automatic processing of threat facial signals. *J. Neurosci., 23*, 5627–5633.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411*, 305–309.

Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge, UK: Cambridge University Press.

Belova, M. A., Paton, J. J., Morrison, S. E., & Salzman, C. D. (2007). Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron, 55*, 970–984.

Bentley, P., Vuilleumier, P., Thiel, C. M., Driver, J., & Dolan, R. J. (2003). Cholinergic enhancement modulates neural correlates of selective attention and emotional processing. *NeuroImage*, *20*, 58–70.

Bishop, S. J., Duncan, J., Brett, M., & Lawrence, A. D. (2004). Prefrontal cortical function and anxiety: Controlling attention to threat-related stimuli. *Nat. Neurosci.*, *7*, 184–188.

Bishop, S. J., Duncan, J., & Lawrence, A. D. (2004). State anxiety modulation of the amygdala response to unattended threat-related stimuli. *J. Neurosci.*, *24*, 10364–10368.

Bishop, S. J., Jenkins, R., & Lawrence, A. D. (2007). Neural processing of fearful faces: Effects of anxiety are gated by perceptual capacity limitations. *Cereb. Cortex*, *17*, 1595–1603.

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2008). Behold the voice of wrath: Cross-modal modulation of visual attention by anger prosody. *Cognition*, *106*, 1497–1503.

Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond fear: Rapid spatial orienting towards positive emotional stimuli. *Psychol. Sci.*, *19*, 362–370.

Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye . . . : Attention capture by infant faces. *Emotion*, *7*, 685–689.

Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Rev.*, *36*, 96–107.

Canli, T., Sivers, H., Whitfield, S. L., Gotlib, I. H., & Gabrieli, J. D. (2002). Amygdala response to happy faces as a function of extraversion. *Science*, *296*, 2191.

Cavada, C., Company, T., Tejedor, J., Cruz-Rizzolo, R. J., & Reinoso-Suarez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex: A review. *Cereb. Cortex*, *10*, 220–242.

Cunningham, W. A., Van Bavel, J. J., & Johnsen, I. R. (2008). Affective flexibility: Evaluative processing goals shape amygdala activity. *Psych. Sci.*, *19*, 152–160.

Dolcos, F., & McCarthy, G. (2006). Brain systems mediating cognitive interference by emotional distraction. *J. Neurosci.*, *26*, 2072–2079.

Driver, J., & Frackowiak, R. (2001). Neurobiological measures of human selective attention. *Neuropsychologia*, 39, 1257–1262.

Driver, J., Vuilleumier, P., & Husain, M. (2004). Spatial neglect and extinction. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences III.* Cambridge, MA: MIT Press.

Fichtenholtz, H. M., Dean, H. L., Dillon, D. G., Yamasaki, H., McCarthy, G., & LaBar, K. S. (2004). Emotion-attention network interactions during a visual oddball task. *Cogn. Brain Res.*, *20*, 67–80.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol. [Hum Percept.]*, *18*, 1030–1044.

Fox, E., Russo, R., & Dutton, K. (2002). Attentional bias for threat: Evidence for delayed disengagement from emotional faces. *Cogn. Emot.*, *16*, 355–379.

Grandjean, D., Sander, D., Lucas, N., Scherer, K. R., & Vuilleumier, P. (2008). Effects of emotional prosody on auditory extinction for voices in patients with spatial neglect. *Neuropsychologia*, *46*, 487–496.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nat. Neurosci.*, *8*, 145–146.

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychol. Sci.*, *14*, 39–46.

Jiang, Y., & He, S. (2006). Cortical responses to invisible faces: Dissociating subsystems for facial-information processing. *Curr. Biol.*, *16*, 2023–2029.

Keil, A., Moratti, S., Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2005). Additive effects of emotional content and spatial selective attention on electrocortical facilitation. *Cereb. Cortex*, *15*, 1187–1197.

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., et al. (2004). Contextual modulation of amygdala responsivity to surprised faces. *J. Cogn. Neurosci.*, *16*, 1730–1745.

Koster, E., Crombez, G., Verschuere, B., & De Houwer, J. (2004). Selective attention to threat in the dot probe paradigm: Differentiating vigilance and difficulty to disengage. *Behav. Res. Ther.*, *42*, 1183–1192.

LeDoux, J. E. (1996). *The emotional brain*. New York: Simon & Shuster.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psych. Bull.*, *132*, 297–326.

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc. Natl. Acad. Sci. USA*, *96*, 1680–1685.

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology*, *30*, 953–958.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *J. Exp. Psychol. [Gen.]*, *130*, 466–478.

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biol. Psychol.*, *77*, 247–265.

Peelen, M., Atkinson, A., Andersson, F., & Vuilleumier, P. (2007). Emotional modulation of body-selective visual areas. *Soc. Cogn. Affect Neurosci.*, *2*, 274–283.

Pegna, A. J., Khateb, A., Lazeyras, F., & Seghier, M. L. (2005). Discriminating emotional faces without primary visual cortices involves the right amygdala. *Nat. Neurosci.*, *8*, 24–25.

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Curr. Opin. Neurobiol.*, *15*, 188–196.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neural and emotional stimuli. *Brain Res. Cogn. Brain Res.*, *15*, 31–45.

Pessoa, L., Padmala, S., & Morland, T. (2005). Fate of unattended fearful faces in the amygdala is determined by both attentional resources and cognitive modulation. *NeuroImage*, *28*, 249–255.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychol. Sci.*, *17*, 292–299.

Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cereb. Cortex*, *14*, 619–633.

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2006). Neural systems for orienting attention to the location of threat signals: An event-related fMRI study. *NeuroImage*, *31*, 920–933.

Pourtois, G., Thut, G., Grave de Peralta, R., Michel, C., & Vuilleumier, P. (2005). Two electrophysiological stages of spatial orienting towards fearful faces: Early temporo-parietal activation preceding gain control in extrastriate visual cortex. *NeuroImage*, *26*, 149–163.

Purkis, H. M., & Lipp, O. V. (2007). Automatic attention does not equal automatic fear: Preferential attention without implicit valence. *Emotion*, *7*, 314–323.

Roesch, M., & Schoenbaum, G. (2006). From associations to expectancies: Orbitofrontal cortex as a gateway between the limbic system and representational memory. In D. H. Zald & S. L. Rauch (Eds.), *The orbitofrontal cortex* (pp. 199–235). Oxford, UK: Oxford University Press.

Rudrauf, D., David, O., Lachaux, J. P., Kovach, C. K., Martinerie, J., Renault, B., et al. (2008). Rapid interactions between the ventral visual stream and emotion-related structures rely on a two-pathway architecture. *J. Neurosci.*, *28*, 2793–2803.

Sabatinelli, D., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2005). Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. *NeuroImage*, *24*, 1265–1270.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Rev. Neurosci.*, *14*, 303–316.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.

Silvert, L., Lepsien, J., Fragopanagos, N., Goolsby, B., Kiss, M., Taylor, J. G., et al. (2007). Influence of attentional demands on the processing of emotional facial expressions in the amygdala. *NeuroImage*, *38*, 357–366.

Smith, N. K., Larsen, J. T., Chartrand, T. L., Cacioppo, J. T., Katafiasz, H. A., & Moran, K. E. (2006). Being bad isn't always good: Affective context moderates the attention bias toward negative information. *J. Pers. Soc. Psychol.*, *90*, 210–220.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends Cogn. Sci.*, *9*, 585–594.

Vuilleumier, P., Armony, J. L., Clarke, K., Husain, M., Driver, J., & Dolan, R. J. (2002). Neural response to emotional faces with and without awareness: Event-related fMRI in a parietal patient with visual extinction and spatial neglect. *Neuropsychologia*, *40*, 2156–2166.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron*, *30*, 829–841.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat. Neurosci.*, *6*, 624–631.

Vuilleumier, P., & Driver, J. (2007). Modulation of visual processing by attention and emotion: Windows on causal interactions between human brain regions. *Philos. Trans. R. Soc. Lond. [Biol.]*, *362*, 837–855.

Vuilleumier, P., & Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia*, *45*, 174–194.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nat. Neurosci.*, *7*, 1271–1278.

Vuilleumier, P., & Schwartz, S. (2001). Beware and be aware: Capture of spatial attention by fear-related stimuli in neglect. *NeuroReport*, *12*, 1119–1122.

Weinberger, N. M. (1998). Tuning the brain by learning and by stimulation of the nucleus basalis. *Trends Cogn. Sci.*, *2*, 271–273.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J. Neurosci.*, *18*, 411–418.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychol. Bull.*, *120*, 3–24.

Williams, M. A., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2005). Differential amygdala responses to happy and fearful facial expressions depend on selective attention. *NeuroImage*, *24*, 417–425.

63 Context Effects and the Amygdala

PAUL J. WHALEN AND F. CAROLINE DAVIS

ABSTRACT The amygdala is a critical component of the brain network responsible for learning about cues that predict biologically relevant outcomes. But these cues are always encountered within a particular context, and this contextual information can modulate amygdala responses. Here, we review neuroimaging studies examining human amygdala responses to specific predictive cues and the modulation of these responses by contextual information. Specifically, we focus on human neuroimaging studies of facial expressions as an example of predictive environmental cues. These studies suggest that regions of the prefrontal cortex and hippocampus are responsive to contextual information and that these responses presumably influence amygdala responses to faces when they are encountered within a given context. These data are interpreted as consistent with the nonhuman animal conditioning literature that supports a role for these regions in facilitating context conditioning.

All environmental events eventually derive their meaning from their context. For example, a person crying at a funeral means something quite different from the very same person crying on a gold medal podium (see Russell & Fernandez-Dols, 1997). Thus, a particular environmental event will give rise to a differential of potential meanings; the context then allows for the selection of the most appropriate meaning of this event within a given circumstance. In this chapter, we will focus on human neuroimaging studies of the amygdala and its interest in events that predict biologically relevant outcomes (e.g., the facial expressions of others). We will see that through its reciprocal connections with the prefrontal cortex, the amygdala can show great flexibility in its responses to predictive cues when informed about contextual considerations. We will then connect this information with the nonhuman animal conditioning literature depicting the known role of the amygdala in facilitating context conditioning.

Using facial expressions to study the human amygdala

Facial expressions mediate a critical portion of our nonverbal communication. From the expressions of others, we can glean information about their internal emotional state, their intentions, and/or their reaction to contextual events in our immediate environment. Facial expressions of emotion have predicted important events for us in the past, and we can use this previous experience to respond appropriately to expressions as we perceive them. In this way, facial expressions can be considered conditioned stimuli.

PAUL J. WHALEN and F. CAROLINE DAVIS Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire

Human Amygdala Responses to Fearful Facial Expressions Are Related to Their Greater Context Dependence

On the basis of animal research showing the importance of the amygdala in fear conditioning (Davis, 1992; Kapp, Whalen, Supple, & Pascoe, 1992; LeDoux, 2000), and studies of bilateral amygdala lesion patients showing deficits in processing the facial expression of fear (Adolphs, Tranel, Damasio, & Damasio, 1994; Broks et al., 1998), early human neuroimaging investigations of amygdala responses to facial expressions focused on fearful expressions (Breiter et al., 1996; Morris et al., 1996; M. L. Phillips et al., 1997; Whalen et al., 1998). Subsequent studies have replicated robust amygdala activations to fearful faces (Pessoa, Japee, Sturman, & Ungerleider, 2004; Whalen et al., 1998; Whalen, Shin, McInerney, & Fischer, 2001). These findings have been presented as consistent with the traditional role of the amygdala in processing threat-related information exclusively.

But additional data from these same amygdala lesion patients suggest that the amygdala is not responding to fearful faces solely because of their threatening nature. (Adolphs et al., 1994). We took this finding as a clue and speculated that perhaps the amygdala is particularly important for processing expressions that predict uncertain (fear) vs. certain (anger) threat. To elaborate, though angry and fearful faces can be equated in terms of the subjective arousal and valence ratings that subjects ascribe to them, they offer very different predictive information value. While both expressions suggest to a viewer that the probability of threat in the immediate context has increased, fearful faces provide no information about the source of that threat, while angry faces embody the very threat that is to be dealt with. Said another way, fearful faces are much more *context dependent* than angry faces are. Fearful faces call for diffuse attention to be directed toward the immediate context ("What is she afraid of?" and "Should I be afraid too?"), while angry faces call for direct attention aimed at the face ("What will he do next?" and "How will I respond?").

In partial support of this hypothesis, studies have shown that fearful expressions are related to greater contextual

monitoring. For example, subjects are better at a perceptual discrimination task on trials that are contextually bound to fearful faces (Phelps, Ling, & Carrasco, 2006). Interestingly, it is known that we, to some degree, mimic the expressions of others (Hess, Blairy, & Philpott, 1999; Dimberg, 1990). It follows then that another person's fearful expression would lead to mimicry of a similar expression on our part. If the detection of fearful expressions leads to facilitated environmental monitoring, then this mimicry could be an important component of this process. Consistent with this notion, recent data show that when subjects make fearful expressions, they show facilitated monitoring of the context (Susskind et al., 2008).

Human neuroimaging data show that amygdala responses may be a critical component of the process that instigates greater environmental monitoring in response to fearful expressions. We have published one study showing that while one portion of the amygdaloid complex appears to track the valence of presented facial expressions, another shows responses that appear to be sensitive to the uncertainty associated with their predicted outcomes (Whalen et al., 2001). Specifically, a ventral portion of the amygdala showed responses of equal magnitude to fearful and angry faces when compared to neutral faces, consistent with the notion that this region of the amygdala may have been responding to the equal negative valence of these expressions. In parallel, a more dorsal portion of the human amygdaloid complex extending into the ventral basal forebrain showed greater responses to fearful faces when directly compared to angry faces. We hypothesized that these dorsal amygdala/ventral forebrain responses may be in reaction to the uncertain information value and greater contextual dependence associated with fearful expressions compared to angry expressions (Whalen et al., 2001).

These data suggest that in addition to threat detection, the amygdala is also recruited when the nature of predicted threat is unclear. Thus the amygdala will be the most indispensable in situations when the organism has the most to learn. If some portion of signal changes within the amygdaloid complex in response to fearful expressions are truly related to the ambiguity of the contextual event that elicited the expression, rather than the negative valence of the face itself, then a compelling demonstration would involve showing a similar amygdala activation to a facial expression that has a similar ambiguity of contextual source but that is not necessarily negatively valenced.

The Facial Expression of Surprise Surprised expressions provide an important comparison for fear. Though neither expression (fear nor surprise) indicates the exact nature of its eliciting event, fearful expressions do provide additional information concerning predicted negative valence. Surprise, on the other hand, can be interpreted either positively or negatively (Tomkins & McCarter, 1964). For example, when asked to provide a story about what is happening to the person pictured showing a surprised expression, some people describe an oncoming car (negative), while others describe an unexpected birthday party (positive) (Davis & Whalen, unpublished observations). Thus surprised facial expressions can be used to (1) reveal important individual differences in the propensity to subjectively ascribe positive or negative valence to an ambiguous facial expression and (2) determine the relationship between these subjective ratings and fMRI signal changes in the amygdala. Further, given their potential dual valence representation, we expected these expressions to be particularly sensitive to context manipulations.

For our first study, we simply had subjects passively view repeated presentations of individuals displaying surprised facial expressions and then had subjects rate the expressions on a scale from 1 (very positive) to 9 (very negative) following the fMRI scanning session. Figure 63.1*A* shows that a lateral ventral region of the amygdala tracked individual differences in valence interpretations of surprised faces. Subjects who interpreted the surprised faces negatively showed higher signal values that correlated with the intensity of these ratings, while subjects who interpreted the faces positively showed lower correlated signal values. Consistent with the assertion made above that the amygdala is also responsive to the ambiguity of contextual source (which exists whether the face is interpreted negatively or positively), figure 63.1*B* shows homogeneous signal increases across all subjects within more dorsal as well as medial portions of the amygdala, despite these differences of opinion related to the valence of the faces. Thus different portions of the amygdaloid complex appear to work on different parts of the problem: predictive aspects of an expression that appear clear to the viewer (e.g., valence—"that face looks negative to me") and, simultaneously, other aspects that remain unclear (e.g., context—"I wonder what she is reacting to").

It might seem a bit surprising that the ventral amygdala did not respond to the surprised faces in some individuals (i.e., those who interpreted the faces positively). Given that one function of the amygdala is to monitor the environment for potential threat, one would have thought that the ventral amygdala would have responded to the potential negativity of the surprised faces similarly in all subjects. Individual differences of this type suggested to us that another region of the brain might be exerting a regulatory influence over the amygdala. Accordingly, we looked to the medial prefrontal cortex (mPFC), given the known reciprocal connections of this region with the amygdala in both the rat and nonhuman primate (e.g., Ghashghaei, Hilgetag, & Barbas, 2006; Stefanacci & Amaral, 2002).

We identified two regions of the mPFC that were correlated with subjects' valence interpretations of surprised faces

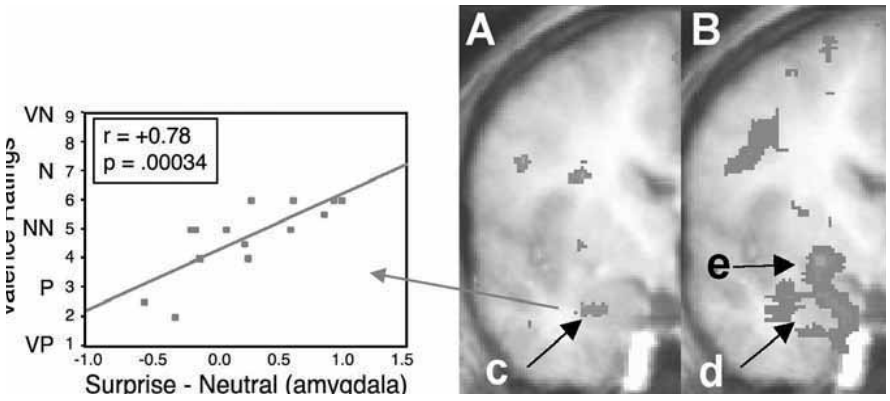


FIGURE 63.1 (*A*) The activated voxels depict the location within the ventral amygdala where a positive correlation with ratings of surprise was observed (arrow c) from the results of H. Kim and colleagues (2003). Scatterplot to the left presents these data. The *x*-axis presents fMRI responses to surprised versus neutral faces, while the *y*-axis presents the valence scale from 1 to 9. Labels on the *y*-axis: VN, very negative; N, negative; NN, neither negative nor positive; P, positive; VP, very positive. (*B*) Voxels at this same anterior-posterior level ($y = -3$) showing a significant main effect for surprised versus neutral faces across all subjects. The maximally activated voxel for this main effect is located in the dorsal amygdala/SI (arrow e). Note that these voxels do not include the voxels in which we observed the significant correlation based upon individual differences (arrow d). Image *B* is thresholded liberally ($p < .05$, uncorrected) to make the points that (1) no trend toward a main effect for surprise existed in the voxels presented in *A*, and (2) we have signal coverage across the entirety of the amygdaloid complex.

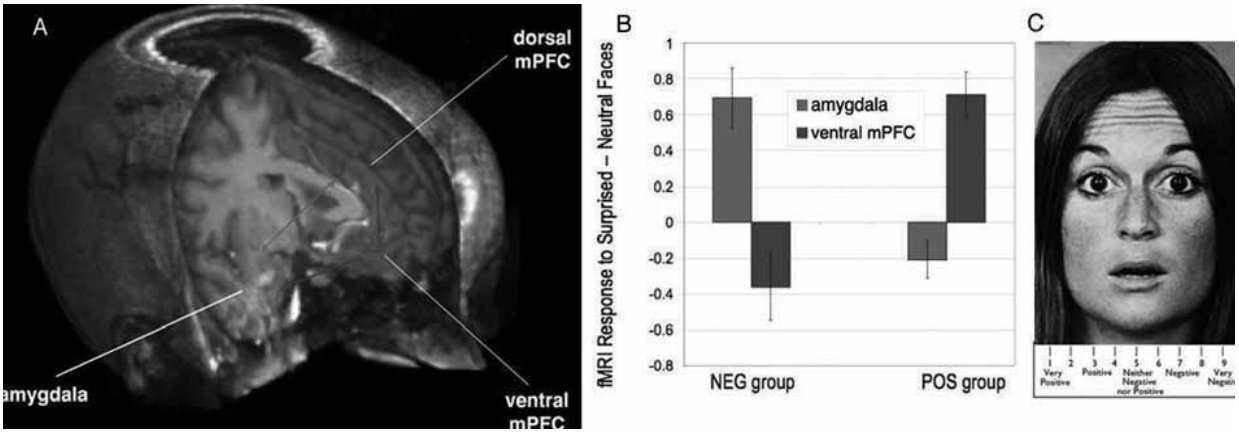


FIGURE 63.2 (*A*) A three-dimensional depiction of the correlational results of H. Kim and colleagues (2003). Amygdala and dorsal mPFC loci that showed a positive correlation with valence ratings of surprise (colored in orange) are also positively correlated with one another (red arrow, $r = +.66$). The ventral mPFC locus that showed a negative correlation with valence ratings of surprise (colored in blue) is also negatively correlated with amygdala (blue arrow; $r = -.69$) and the dorsal mPFC (blue arrow; $r = -.62$). (*B*) Bar graph focusing on the inverse relationship between the amygdala and ventral mPFC in subjects who interepted the surprised faces either positively (POS) or negatively (NEG). (*C*) An example of the surprised faces and valence scale used to rate them. (See color plate 79.)

(see figure 63.2). Like the amygdala, a dorsal region of the mPFC (specifically the rostral ACC; see H. Kim et al., 2003) displayed a positive relationship with negative valence ratings (i.e., higher activity with more negative ratings). A ventral region of the mPFC (ventral ACC; see H. Kim, Somerville, Johnstone, Alexander, & Whalen, 2003) showed an opposite relationship with valence ratings of surprised faces compared to the amygdala and dorsal mPFC (i.e., higher activity with more positive ratings). That this ventral mPFC activity was inversely correlated with the *ventral* amygdala, is consistent with the location of known prefrontal-amygdala connections in nonhuman primates (Ghashghaei et al., 2006).

Figure 63.2*A* presents a three-dimensional representation of these regions where activity in the ventral mPFC is inversely related to activity in the amygdala and dorsal mPFC. Figure 63.2*B* presents a bar graph that focuses on the inverse relationship between the ventral mPFC and the amygdala, showing that subjects who interpreted these faces negatively showed high amygdala responses and correlated low ventral mPFC responses, while the subjects who

interpreted them positively showed high ventral mPFC responses and correlated low amygdala responses. Note that this inversely correlated ventral mPFC-amygdala fMRI activity was observed during passive viewing. That is, activity measured while subjects viewed repeating surprised faces predicted the valence ratings the subjects assigned to these faces following the scanning session. These data suggest that these brain activations related to valence calculations were relatively automatic/implicit while subjects passively viewed the faces (H. Kim et al., 2003).

One interpretation of these data is that in response to ambiguous surprised expressions, a regulatory override message from the ventral region of the mPFC is required to interpret these faces as positively valenced. Inherent in this assertion is the presumption that the amygdala is involved in an initial default negative interpretation of surprised faces in all subjects, after which some subjects are able to regulate this amygdala response and respond more positively. More generally, these data suggest that surprised faces constitute a simple means for tapping into a purported medial prefrontal-amygdala regulatory circuit, a circuit that mediates a similar override function during the extinction of conditioned fear responses (Milad & Quirk, 2002; Morgan, Romanski, & LeDoux, 1993; Phelps, Delgado, Nearing, & LeDoux, 2004; see Oler, Quirk, & Whalen, in press, for further discussion).

Modulating Amygdala Responses to Surprised Faces by Modulating the Context in Which They Are Presented If the hypothesis that greater amygdala activity is observed to fearful and surprised faces because they are more context-dependent is sound, then manipulation of contextual information during presentations of surprised faces should differentially affect amygdala activation. Figure 63.3*A* shows the design of an experiment in which participants were explicitly told whether the presented surprised face was in reaction to a negative event or a positive event. Instead of letting subjects decide for themselves whether a particular surprised face was negative or positive, each surprised face was preceded by a contextual sentence describing either a negative or a positive event (e.g., "He just lost $500" or "He just found $500"). Subjects were told that the expression that followed the sentence was in direct reaction to the described event. Figure 63.3*B* shows that the ventral amygdala did not show differential responses to the sentences themselves but did show differential responses to the faces that were primed by the contextual sentences. Specifically, although the presented contextual sentences were carefully matched for ratings of subjective arousal and intensity of valence, amygdala responses were greater to the surprised faces that were presented in a negative context compared to a positive context (H. Kim et al., 2004). Note the lack of activation across more dorsal portions of the amygdaloid complex (see figure 63.1); we suggest that this is due to the fact that the sentences disambiguated the eliciting source of the surprised faces; thus only the activation based upon valence, not the main effect for surprise, is observed. Also, consistent with the findings of H. Kim and colleagues (2003), a region of the mPFC (dorsal ACC) showed activation to negative and positive contextual sentences (see figure 63.4).

We speculate that across the two studies, the mPFC exerted an influence over amygdala activity leading to lower amygdala signal intensities when surprised faces were interpreted positively. In the latter study (H. Kim et al., 2004),

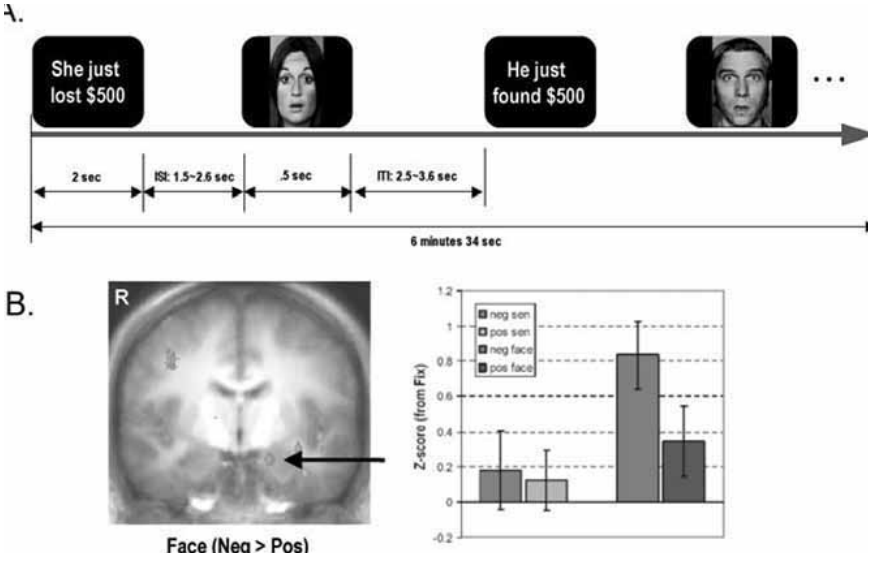


Figure 63.3 Paradigm from H. Kim and colleagues (2004).

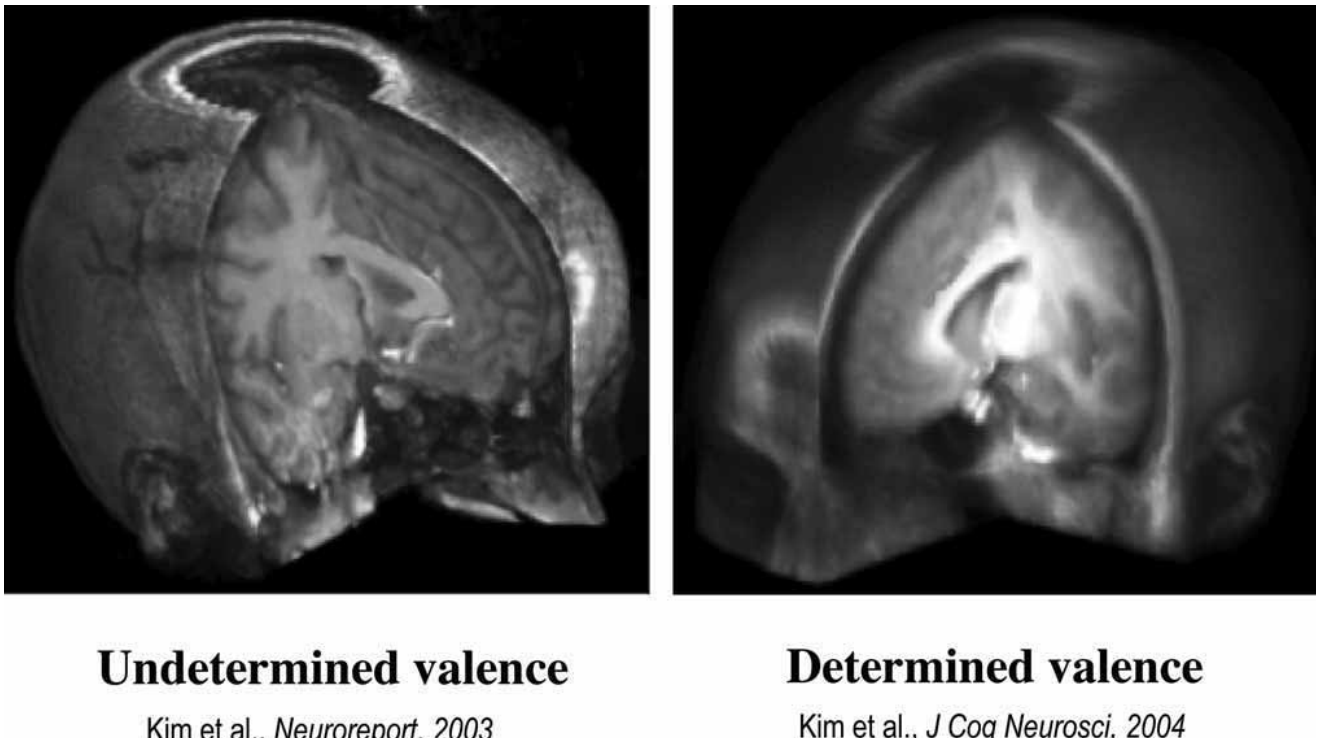


FIGURE 63.4 Comparison of results of H. Kim and colleagues (2003) with those of H. Kim and colleagues (2004). (See color plate 80.)

the basis of this regulation was derived from the contextual sentences, while in the former study (H. Kim et al., 2003), this information was likely derived from past experience/reinforcement history with these expressions. Note the striking laterality difference observed between these two studies where a right-sided prefrontal-amygdala network processed surprised faces whose valence was left undetermined (H. Kim et al., 2003), while a left-sided network processed surprised faces whose valence was determined (see H. Kim et al., 2004, for discussion). Note also the lack of involvement of the vmPFC when the sentences disambiguated the faces. Perhaps a regulatory input from the vmPFC is only necessary when subjects have to discern the potential valence of the expressions for themselves (see H. Kim et al., 2004, for discussion). Alternatively, numerous studies have implicated the vmPFC in self-relevant processing (e.g., Moran, Macrae, Heatherton, Wyland, & Kelley, 2006). In H. Kim and colleagues' (2004) study, the sentences provided information about the person pictured. One possibility is that when no such contextual information is available, we default to a more self-relevant hypothesis testing mode ("I wonder what that expression means for me"), a mode that involves the vmPFC. Future studies will be necessary to sort out these possibilities. For now, these initial studies have delineated other brain structures with which the amygdala interacts to help disambiguate the valence of ambiguous facial expressions of emotion.

To summarize, facial expressions that are ambiguous are highly susceptible to contextual manipulation. The amygdala in concert with the prefrontal cortex uses contextual information to resolve this ambiguity (H. Kim et al., 2004). These neuroimaging data in humans using facial expressions as the predictive cues of interest are consistent with the conditioning literature largely carried out by using nonhuman animal subjects. These studies in nonhuman animals suggest that the amygdala is involved in processing contextual information through interactions with other brain regions, such as the hippocampus and prefrontal cortex. Below, we describe recent human neuroimaging studies of context conditioning that support a similar role for the amygdala, hippocampus, and prefrontal cortex, while drawing parallels with the facial expression work we have just described.

Context conditioning

Similar to the above discussion of the contextual modulation of facial expressions, the vast Pavlovian conditioning literature demonstrates the critical role that context plays in determining the meaning of conditioned "cues." The amygdala is consistently implicated in "cue conditioning," in which an animal learns that a distinct cue such as a light or tone (conditioned stimulus, CS) predicts a biologically relevant event such as a food reward or painful shock (unconditioned stimulus, US; see, e.g., LeDoux, 1996; Kapp et al., 1992).

However, the animal always learns such associations in a particular context. Though the processing of contextual configurations associated with the presentation of specific conditioned cues is known to be dependent on the dorsal hippocampus (J. J. Kim & Fanselow, 1992), the amygdala has been shown to be a critical facilitator of such learning (R. G. Phillips & LeDoux, 1992). This contextual information is likely conveyed to the amygdala via multiple neural pathways such as through the hippocampus (Maren, 2001), the orbitofrontal cortex (Stefanacci & Amaral, 2002), and the ventral medial prefrontal cortex (Milad & Quirk, 2002).

Grillion and colleagues recently provided evidence in the human that the amygdala, hippocampus, and prefrontal cortex are recruited during a contextual conditioning task (Alvarez, Biggs, Chen, Pine, & Grillon, 2008). In this neuroimaging experiment (see figure 63.5), subjects learned that a certain virtual reality context predicted shock (CX+) while another context predicted safety (CX–). Brain areas showing greater activity during the CX+ compared to the CX– included the amygdala, prefrontal cortex, and hippocampus. Importantly, the authors also employed structural equation modeling to infer directional relationships between activity in these structures and concluded that the prefrontal cortex and hippocampus might be providing contextual information to the amygdala that then initiates conditioned responding appropriate to a particular context.

When a cue has clear predictive value (i.e., a tone *always* predicts a shock), then it is less dependent upon contextual cues to retrieve the appropriate meaning of that cue. However, there are many situations in which the predictive meaning of a cue is more ambiguous, and in these situations, contextual information becomes very important (similar to the arguments presented above in relation to surprised faces; see Oler et al., in press). A classic illustration of this is extinction. During extinction, an animal learns that a cue that previously had very clear predictive value (a tone always predicts a shock) now has multiple meanings (Bouton, 1994). That is, in some contexts, the tone predicts a shock, whereas in others, it predicts no shock. In this scenario, context plays an important role in influencing the animal's response to a CS, but it is not the direct relationship between the context and the occurrence (or nonoccurrence) of the US that is important. Rather, the context helps to activate the appropriate representation of the CS (Bouton, 1994; Bouton, Westbrook, Corcoran, & Maren, 2006).

In the human, a circuitry involving interactions between the amygdala and the hippocampus, OFC, and vmPFC during contextually dependent extinction learning has been documented. During neuroimaging, Milad and colleagues (2007) trained subjects to expect shocks following the presentation of a light in one context (context A) and extinguished this relationship in the same context. They found increases in activity in both the amygdala and vmPFC during extinction learning. Interestingly, the authors brought subjects back the next day and exposed them to the extinguished CS in either the extinguished context (context A) or a new context (context B). During this session, they observed increases in vmPFC and hippocampal activity to the extinguished context compared to the new context. Critically, functional connectivity analyses suggested that activity in the vmPFC on day 2 was positively correlated with activity in both the hippocampus and the amygdala. Taken together, these studies provide strong support that a neurocircuitry involving the amygdala, hippocampus, and prefrontal cortex supports the ability to learn about the relationship between environmental events that have biologically relevant predictive value and the contexts in which they occur.

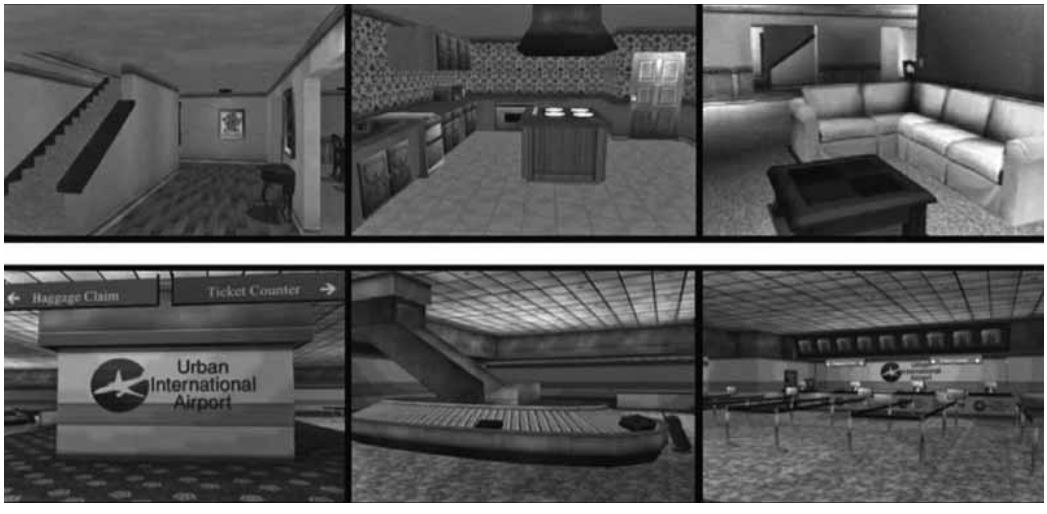


Figure 63.5 Examples of the virtual reality environments used by Alvarez and colleagues (2008) that served as conditioned contexts. (Reproduced with permission.)

Experimental context as a determinant of fMRI activations to specific stimuli

Often, we expect that there will be hard rules by which neural systems operate, for example, "The amygdala responds to negative more than positive stimuli" or "The amygdala is influenced by arousal, not valence." But it would be strange if a neural system that functioned to support biologically relevant learning was so inflexible. It would make more sense if there were some basic operating principles guiding amygdala activity across all species that interact with the more flexible cortical circuitry.

This idea is best kept in mind when one considers that in experimental designs, the stimuli that are included as other conditions of interest, control conditions, or foil conditions effectively comprise the context in which the main stimuli of interest will be considered. These stimuli will not necessarily be independent of the main stimuli but will instead shade the very light in which these stimuli are perceived. Here we offer four examples.

Is Neutral Ever Neutral? Neutral faces often constitute the baseline condition for neuroimaging studies assessing amygdala responses to fearful faces. This is a sound strategy, as any responses observed to fearful faces when directly contrasted with neutral faces are soundly interpreted as in response to the fearful expression itself. But such a design will necessarily ignore any interest the amygdala might have in neutral faces in their own right. An older psychology literature suggests that the interpretation of neutral faces is strongly influenced by the clearer valence of other categorical expressions in the immediate context. That is, neutral expressions are rated as more positive when negative expressions such as fear are in the immediate context, while they are rated as more negative when positive expressions are in the context (Russell & Fehr, 1987). With this in mind, Somerville, Kim, Johnstone, Alexander, and Whalen (2004) presented neutral expressions accompanied only by happy expressions in the experimental context. They observed bilateral amygdala activation in response to the neutral expressions that correlated with individual differences in subjects' level of anxiety, greater levels of anxiety being associated with greater amygdala responses to neutral faces. The implication of this finding for the present discussion is that it might well have been this particular experimental context that created a situation in which the amygdala tracked neutral faces more than it would have otherwise. Future studies that modify the valence of the expressions that make up the context for neutral faces can directly test this hypothesis.

Static Versus Dynamic Another example of the influence of experimental context is derived by comparing results from two separate studies. The amygdala has clearly been shown to respond to static photographs of facial expressions across numerous neuroimaging studies. While interesting, it is obvious that these stimuli lack a degree of ecological validity, since the expressions that we encounter on a daily basis are always dynamic. One study using only static images showed that the amygdala responds to both fearful and angry faces (compared to neutral faces) and that a greater spatial proportion of the amygdala responded to fearful faces (Whalen et al., 2001). A subsequent study also included these static image conditions but did not replicate this effect (LaBar, Crupain, Voyvodic, & McCarthy, 2003). Critically, this study also included dynamic displays of these stimuli within the experimental context. It is probable, then, that in such an experimental context, the amygdala tracked the more salient and interesting dynamic displays to the exclusion of the static displays (LaBar et al., 2003). In the experimental context that was devoid of dynamic displays, the amygdala tracked the best available predictive information (i.e., static displays) (Whalen et al., 2001). Thus, a more flexible rule might be that in any given situation, the amygdala will track the best potential source of information from which to learn. This notion is consistent with other data showing that while the amygdala is responsive to facial expressions of emotion, it responds even more to facial expressions embedded within an associative learning context (Hooker, Germine, Knight, & D'Esposito, 2006). A more recent study demonstrated that amygdala activation is observed when subjects are asked to match a given facial expression to an appropriate context (Sommer, Dohnel, Meinhardt, & Hajak, 2008). Taken together, these data are consistent with the idea that the amygdala is particularly interested in facial expressions because of their predictive value, which will ultimately be gleaned from their current environmental context.

Instructions Task instructions are another example of contextual information that can influence neural processing of presented stimuli. That is, the very same stimuli will be processed in very different ways depending on how subjects are instructed to interpret these stimuli. For example, Anderson, Christoff, Panitz, De Rosa, and Gabrieli (2003) compared amygdala responses to fearful and disgusted faces. Earlier studies had suggested that the amygdala is interested in fearful expressions but not disgusted expressions (e.g., M. L. Phillips et al., 1997). Anderson and colleagues (2003) presented these facial expressions on a screen superimposed with pictures of houses. During some time periods, subjects were instructed to pay attention to the faces, while during others they attended to the houses. Replicating the earlier work, Anderson and colleagues (2003) showed that the amygdala responded to fearful faces whether or not subjects were instructed to pay attention to these expressions. But

the amygdala responded to disgusted faces only when they were not attended to. Thus the amygdala tracked the presence of disgusted faces to a greater degree when they were being implicitly processed. The authors interpreted their data to suggest that the amygdala widens its focus from fear-specific to more generalized threat-specific as attention to faces moves from explicit to implicit. For the purposes of this discussion, we see that task instructions constitute a powerful context that will influence patterns of neural responsivity.

Face as Context Since we have focused on facial expressions of emotion as predictive cues, it is interesting to consider the possibility that the face itself actually constitutes a context, and the muscle movements can be considered the predictive cues. We have enough experience with faces to know where to look within this context to discern predictive information. Numerous studies have shown that the eye and mouth regions are the first place we look for such information. What neural systems support this ability? One hint comes from studies showing that subjects with amygdala damage fail to properly monitor the eye region of the face (Adolphs et al., 2005). Similarly, rats with amygdala damage fail to turn their head to the region of the cage where a predictive light is located (see Gallagher & Holland, 1992). These data support the notion that the amygdala is part of an orienting system that initiates behaviors (e.g., eye scanning, head turning) that facilitate the search for predictive information in locations within the context where the organism has last encountered predictive cues (see Whalen et al., 2009, for discussion). Such behaviors could be initiated by recognizing that one has encountered a "face context" or a "cage context," respectively. These observations may effectively summarize why lesions of the amygdala affect context conditioning. While areas of the brain such as the hippocampus and the prefrontal cortex are assessing the context as a configuration, the amygdala, like a prosopagnosic, might be surveying the individual elements that ultimately constitute that context.

Conclusions

Here, we have presented studies that assess amygdala activation to predictive facial expressions that are particularly context dependent (fear and surprise). We have seen that amygdala responses to fearful and surprised faces can be as much about what these faces clearly tell you ("I predict something negative") as they are about what they do not yet tell you ("I offer no information about what exactly that might be"). For this information, the amygdala interacts with other brain regions such as the hippocampus and prefrontal cortex to access contextual information that might serve to disambiguate the predictive nature of these expressions. The fact that surprised faces are ambiguous with respect to valence (i.e., have predicted both positive and negative events in the past) means that contextual information will be even more important in determining their meaning compared to fear. More generally, we have seen that data in human subjects in response to facial expressions of emotion complement data from nonhuman animals showing a clear role for the amygdala in facilitating contextual processing (R. G. Phillips & LeDoux, 1992). Further, recent neuroimaging studies have delineated a similar neurocircuitry supporting context conditioning in the human.

Finally, we reviewed studies demonstrating that the experimental context (e.g., instructions, other stimuli present) of a given study constitute a powerful context that will influence fMRI activations to individual events within that context. Thus the amygdala will track fearful faces one moment and disgusted faces the next (Anderson et al., 2003). It will track static displays of expressions in one study (Whalen et al., 2001) but ignore them in another experimental context that includes dynamic displays (LaBar et al., 2003). And it can track neutral faces depending upon the nature of adjacent expressions (Somerville et al., 2004). Thus the amygdala will be bit of a chameleon, eluding categorization based upon responses that follow a single rule. Instead, in collaboration with other brain regions, it will be an integral part of a circuit that tracks the predictive meaning of stimuli in light of their contextual implications.

REFERENCES

Adolphs, R., Tranel, D., Damasio, A. R., & Damasio, H. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*, 669–672.

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*, 68–72.

Alvarez, R. P., Biggs, A., Chen, G., Pine, D. S., & Grillon, C. (2008). Contextual fear conditioning in humans: Cortical-hippocampal and amygdala contributions. *J. Neurosci.*, *28*(24), 6211–6219.

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. (2003). Neural correlates of the automatic processing of threat facial signals. *J. Neurosci.*, *23*(13), 5627–5633.

Bouton, M. E. (1994). Context, ambiguity, and classical conditioning. *Curr. Dir. Psychol. Sci.*, *3*(2): 49–53.

Bouton, M. E., Westbrook, F., Corcoran, K. A., & Maren, S. (2006). Contextual and temporal modulation of extinction: Behavioral and biological mechanisms. *Biol. Psychiatry*, *60*, 352–360.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, *17*(5), 875–887.

Broks, P., Young, A. W., Maratos, E. J., Coffey, P. J., Calder, A. J., Isaac, C. L., et al. (1998). Face processing impairments after encephalitis: Amygdala damage and recognition of fear. *Neuropsychologia*, *36*(1), 59–70.

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annu. Rev. Neurosci.*, *15*, 353–375.

Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology*, *27*, 481–494.

Gallagher, M., & Holland, P. (1992). Understanding the function of the central nucleus: Is simple conditioning enough? In J. P. Aggelton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (pp. 307–322). New York: Wiley-Liss.

Ghashghaei, H. T., Hilgetag, C. C., & Barbas, H. (2006). Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *NeuroImage*, *34*, 905–923.

Hess, U., Blairy, S., & Philipott, P. (1999). Facial mimicry. In P. Philipott, R. Feldman, & E. Coates (Eds.), *The social context of nonverbal behavior* (pp. 213–241). New York: Cambridge University Press.

Hooker, C. I., Germine, L. T., Knight, R. T., & D'Esposito, M. (2006). Amygdala response to facial expressions reflects emotional learning. *J. Neurosci.*, *26*(35), 8915–8922.

Kapp, B. S., Whalen, P. J., Supple, W. F., & Pascoe, J. P. (1992). Amygdaloid contributions to conditioned arousal and sensory information processing. In J. P. Aggelton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (pp. 229–254). New York: Wiley-Liss.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *NeuroReport*, *14*(18), 2317–2322.

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., et al. (2004). Contextual modulation of amygdala responsivity to surprised faces. *J. Cogn. Neurosci.*, *16*(10), 1730–1745.

Kim, J. J., & Fanselow, M. S. (1992). Modality-specific retrograde amnesia of fear. *Science*, *256*, 675–677.

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cereb. Cortex*, *13*(10), 1023–1033.

LeDoux, J. E. (1996). *The emotional brain. The mysterious underpinnings of emotional life*. New York: Simon & Schuster.

LeDoux, J. E. (2000). The amygdala and emotion: A view through fear. In J. P. Aggleton (Ed.), *The amygdala: A functional analysis* (2nd ed., pp. 289–310). New York: Oxford University Press.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.*, *24*, 897–931.

Milad, M. R., & Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*, *420*, 70–74.

Milad, M. R., Wright, C. I., Orr, S. P., Pitman, R. K., Quirk, G. J., & Rauch, S. L. (2007). Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol. Psychiatry*, *62*(5), 446–454.

Moran, J. M., Macrae, C. N., Heatherton, T. F., Wyland, C. L., & Kelley, W. M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *J. Cogn. Neurosci.*, *18*, 1586–94.

Morgan, M. A., Romanski, L. M., & LeDoux, J. E. (1993). Extinction of emotional learning: Contribution of medial prefrontal cortex. *Neurosci. Lett.*, *163*, 109–113.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, *383*, 812–814.

Oler, J. A., Quirk, G. J., & Whalen, P. J. (in press). Cingulo-amygdala interactions in surprise and extinction: Interpreting associative ambiguity. In B. Vogt (Ed.), *Cingulate neurobiology and disease: Vol. 1. Infrastructure, diagnosis, and treatment.* New York: Oxford University Press.

Pessoa, L., Japee, S., Sturman, D., & Ungerleider, L. G. (2006). Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cereb. Cortex*, *16*(3), 366–375.

Phelps, E. A., Delgado, M. R., Nearing, K. I., & LeDoux, J. E. (2004). Extinction learning in humans: Role of the amygdala and vmPFC. *Neuron*, *43*, 897–905.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychol. Sci.*, *17*, 292–299.

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*(6650), 495–498.

Phillips, R. G., & LeDoux, J. E. (1992). Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behav. Neurosci.*, *106*(2), 274–285.

Russell, J. A., & Fernandez-Dols, J. M. (1997). *The psychology of facial expression.* Cambridge, UK: Cambridge University Press.

Russell, J. A., & Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *J. Exp. Psychol.*, *116*(3), 223–237.

Somerville, L. H., Kim, H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2004). Human amygdala responses during presentation of happy and neutral faces: Correlations with state anxiety. *Biol. Psychiatry*, *55*(9), 897–903.

Sommer, M. Dohnel, K., Meinhardt, K., & Hajak, G. (2008). Decoding of affective facial expressions in the context of emotional situations. *Neuropsychologia*, *46*(11), 2615–2621.

Stefanacci, L., & Amaral, D. G. (2002). Some observations of cortical inputs to the macaque monkey amygdala: An anterograde tracing study. *J. Comp. Neurol.*, *451*(4), 301–323.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nat. Neurosci.*, *11*(7), 843–850.

Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Percept. Mot. Skills*, *18*, 119–158.

Whalen, P. J., Davis, F. C., Oler, J. A., Kim, H., Kim, M. J., & Neta, M. (2009). The human amygdala and facial expressions of emotion. In E. A. Phelps & P. J. Whalen (Eds.), *The human amygdala* (pp. 265–288). New York: Guilford.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J. Neurosci.*, *18*(1), 411–418.

Whalen, P. J., Shin, L. M., McInerney, S. C., & Fischer, H. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion*, *1*(1), 70–83.

64 Neurogenetic Studies of Variability in Human Emotion

AHMAD R. HARIRI

ABSTRACT As research in cognitive neuroscience has progressed in the last decades, there have been many important technological and methodological advances in the increasingly complimentary fields of molecular genetics and neuroimaging. These advances have facilitated fruitful collaboration across once disparate disciplines, with early results shedding new light on the mechanisms that give rise to individual differences in complex behaviors and related psychiatric disorders. At the leading edge of such efforts is imaging genetics, an experimental strategy for the effective integration of molecular genetics and neuroimaging technologies for the study of biological mechanisms that mediate individual differences in behavior and related risk for psychiatric disorders. Imaging genetics studies have provided a more complex and nuanced understanding of the pathways and mechanisms through which the dynamic interplay of genes, brain, and environment shapes variability in behavior. This chapter provides a brief overview of these studies and discusses the broader potential of imaging genetics, through its orchestrated application with studies of environmental effects and its continued integration with basic animal research, to inform risk and resiliency.

Conceptual basis

Individual differences in trait affect, personality, and temperament are important predictors of vulnerability to neuropsychiatric disorders including depression, anxiety, and addiction. Accordingly, identifying the biological mechanisms that give rise to trait individual differences affords unique opportunity to develop both predictive markers of disease liability and identify novel targets for individualized treatment. In the past five years, human neuroimaging studies, especially those employing blood oxygen level dependent functional magnetic resonance imaging (BOLD fMRI), have begun to reveal the neural substrates of interindividual variability in these and related constructs (Bertolino et al., 2005; Brown, Manuck, Flory, & Hariri, 2006; Haas, Omura, Constable, & Canli, 2007; Hariri et al., 2006; Pezawas et al., 2005; Somerville, Kim, Johnstone, Alexander, & Whalen, 2004). Moreover, recent studies have established that BOLD fMRI measures represent temporally stable and reliable indices of brain function (Johnstone et al., 2005; Manuck, Brown, Forbes, & Hariri, 2007). Thus, much like their behavioral counterparts, patterns of brain activation represent enduring, traitlike phenomena, which in and of themselves may serve as important markers of liability and pathophysiology.

As neuroimaging studies continue to illustrate the predictive relationship between regional brain activation and traitlike behaviors (e.g., increased amygdala reactivity predicts core features of anxious temperament), an important next step is to systematically identify the underlying mechanisms that drive variability in brain circuit function. In this regard, recent neuroimaging studies employing pharmacological challenge paradigms, principally targeting monoamine neurotransmission, have revealed that even subtle alterations in dopaminergic, noradrenergic, and serotonergic signaling can have a profound impact on the functional response of brain circuitries that support affect, personality, and temperament (Bigos et al., 2008; De Martino, Strange, & Dolan, 2008; Hariri, Mattay, Tessitore, Fera, et al., 2002; Harmer, Mackay, Reid, Cowen, & Goodwin, 2006). Similarly, multimodal neuroimaging approaches have provided evidence for directionally specific relationships between key components of monoaminergic signaling cascades, assessed with radiotracer position emission tomography (PET), and brain function, assessed with BOLD fMRI (Fisher, Meltzer, Ziolko, Price, & Hariri, 2006; Rhodes et al., 2007). Collectively, pharmacological challenge neuroimaging and multimodal PET/fMRI are revealing how variability in behaviorally relevant brain activation emerges as a function of underlying variability in key brain neurotransmission systems (e.g., increased serotonin signaling predicting increased amygdala reactivity).

The next logical step is to identify the sources of interindividual variability in these key neurochemical signaling mechanisms. In the modern era of human molecular genetics, this step is firmly planted in the direction of identifying the relationships between common variation in the genes encoding components of these signaling cascades, their protein products, and, subsequently, brain circuit function. As sequence variation across individuals represents the ultimate wellspring of variability in emergent neurobiological

AHMAD R. HARIRI Department of Psychology and Neuroscience, Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina

and related behavioral processes, understanding the relationships between genes, brain, and behavior is critical for establishing the etiology and pathophysiology of psychiatric disease. The emerging field of imaging genetics seeks to establish a principled framework for the integration of modern molecular genetics and neuroimaging technologies toward the ultimate goal of identifying truly predictive makers of disease vulnerability (Hariri & Weinberger, 2003a).

Why study genes?

Genes have an unparalleled potential impact on all levels of biology. In the context of disease states, particularly behavioral disorders, genes are fundamental to our understanding of the mechanisms involved in the development of disease. Whereas most human behaviors cannot be explained by genes alone, and certainly much of the variance in aspects of brain information processing will not be genetically determined directly, variations in a genetic sequence that affect gene function will contribute a substantial amount of variance to these more complex phenomena. This conclusion is implicit in results garnered from twin studies that have demonstrated a strong genetic contribution to variability (~40–70%) in aspects of cognition, temperament, and personality (Plomin, Owen, & McGuffin, 1994). Similarly, many psychiatric illnesses, which often cluster within families, have a significant genetic basis (Kendler, Prescott, Myers, & Neale, 2003). Genes, therefore, have the potential to identify underlying mechanisms of variability in behavior and disease risk. Within this context, imaging genetics represents the specific ability to understand the neurobiological mechanisms through which genes may impact variability in these emergent phenomena.

The classic approach used in genetic association analyses involves the use of candidate genes. A candidate gene is a gene that is implicated in the manifestation of a particular behavioral or clinical (i.e., disease-related) phenotype through its effects on a biological process implicated in the same phenotype. With this approach, a genetic variation (or polymorphism) that potentially affects the function of a behaviorally or clinically relevant biological process is identified, and then deviations in the frequency of one variant (or allele) in populations expressing the phenotype are determined. Ideally, the genetic variation should have an impact on molecular or cellular function of the gene or protein (i.e., be a functional variation), and the target phenotype should be stable, robust, and quantifiable. Within the imaging genetics framework, the target phenotype is usually a physiological response of the brain during specific behavioral processes (e.g., amygdala reactivity when viewing threatening facial expressions).

Why neuroimaging?

Previous investigations of candidate genes have attempted to associate functional polymorphisms directly with a behavior; however, such findings have been weak and inconsistent, as there are considerable interindividual differences in the dimensions of observed behavior as well as subjectivity in behavioral measures, often requiring daunting sample sizes to detect even small gene effects (Burmeister, McInnis, & Zollner, 2008). More important, gene effects are not expressed directly at the level of behavior but rather are mediated by effects on molecular and cellular cascades biasing information processing in brain circuitries that mediate behavioral responses to environmental challenge. Neuroimaging, therefore, provides an efficient and effective tool with which to explore the functional impact of brain-relevant genetic polymorphisms and identify neural pathways through which these variants contribute to the emergence of variability in behavior and disease risk.

Basic principles of imaging genetics

Three basic principles have been articulated for imaging genetics (Hariri & Weinberger, 2003b): (1) selection of candidate genes, (2) control for nongenotype factors, and (3) task selection. Well-defined functional polymorphisms (single nucleotide polymorphisms (SNPs) or other structural variants) in coding or promoter regions previously linked with specific physiological effects at the cellular level whose impact has been described in distinct brain regions are an ideal starting point. Selecting variants that have known neurobiological consequences (e.g., increases in serotonin signaling) is important because of an emphasis in imaging genetics on specifying mechanisms through which genes affect brain and related behavior. Because potential genetic effects are still relatively small in comparison to typically large effects of age, sex, and IQ, as well as environmental influences (e.g., illness, injury, substance abuse), controlling for these potential confounds is necessary. Furthermore, since imaging genetics studies focus on a single or relatively few polymorphisms against a background of millions, these studies must carefully control for occult differences in genetic background (i.e., ancestry or population stratification) of the genotype groups. Because of small genetic effects, the choice of a probe of functional brain circuitry is imperative; therefore, a well-characterized behavioral paradigm should be employed to maximize sensitivity and inferential value. The ideal tasks for these investigations are thus ones that have been established to engage specific brain systems robustly in all individuals as well as displaying variance across individuals.

Imaging genetics studies of amygdala reactivity and human emotion

The following sections provide introductory examples of how imaging genetics can lead to insights about the biological mechanisms that underlie individual differences in complex behavioral traits. In many cases, the genetically driven variability in behavior represents an intermediate phenotype that confers risk for neuropsychiatric disease in the context of specific environmental influences. In each group of studies, neuroimaging was employed as a complementary approach to identify effects of genetic polymorphisms, often resulting in molecular and cellular alterations, as well as being associated with specific behaviors and/or disease states, of discrete neural circuitries that support the associated behavioral or clinical phenomena. In addition, each study incorporated the basic principles of imaging genetics described above, such as the implementation of rigorous controls for nongenetic factors such as age, sex, population stratification, and performance on the experimental task. All of the studies also capitalized on existing functional paradigms designed to explore physiological aspects of distinct neural systems.

Serotonin Transporter (5-HTT) The vast potential of imaging genetics has been most dramatically highlighted in recent studies whose collective results demonstrate that common sequence variation in the human serotonin transporter gene is associated with downstream alterations in serotonin signaling cascades that result in relatively increased serotonin signaling and, eventually, increased amygdala reactivity to environmental threat (Hariri & Holmes, 2006). Abnormal 5-HT neurotransmission has been implicated in the pathophysiology of mood and anxiety disorders and has been a target of pharmacological intervention (e.g., SSRIs). In comparison to the 5-HTTLPR long (L) allele, the short (S) allele has been associated with alterations conferring relatively increased 5-HT signaling (Hariri & Holmes, 2006). At the behavioral level, possession of either one or two copies of the S allele has been associated with increased levels of temperamental anxiety (Munafo, Clark, & Flint, 2005; Schinka, Busch, & Robichaux-Keene, 2004; Sen, Burmeister, & Ghosh, 2004), conditioned fear responses (Garpenstrand, Annas, Ekblom, Oreland, & Fredrikson, 2001), and development of depression, particularly in the context of environmental stress (Caspi et al., 2003; Kendler, Kuhn, Vittum, Prescott, & Riley, 2005).

Against this background, imaging genetics has revealed that threat-related reactivity of the amygdala, a brain region that is critical in mediating behavioral and physiological arousal, is significantly increased in S allele carriers in comparison to L allele homozygotes (Hariri, Mattay, Tessitore, Kolachana, et al., 2002). This pattern already represents one of the most positively replicated findings not only in the nascent field of imaging genetics but also in behavioral and psychiatric genetics (Munafo, Brown, & Hariri, 2008). In addition, the 5-HTTLPR S allele has been further linked with reduced gray matter volume in and functional coupling between the amygdala and medial prefrontal cortex (Pezawas et al., 2005). As the magnitude of threat-related amygdala reactivity (as well as its functional coupling with medial prefrontal cortex) is associated with temperamental anxiety, these imaging genetics findings suggest that the 5-HTTLPR S allele may be associated with increased risk for depression upon exposure to environmental stressors because of the polymorphisms influence on the reactivity of this corticolimbic circuitry. The imaging genetics research with the 5-HTTLPR highlights the effectiveness of this strategy in illuminating specific mechanisms that mediate individual variability in behavior and risk for disease.

Monoamine Oxidase-A To the extent to which the effects of the 5-HTTLPR variant on corticolimbic development and function related to emotion processing are serotonin mediated, it would be expected that other genes related to serotonin function would show similar effects on the function of this circuitry. 5-HT neurotransmission is also regulated through intracellular degradation via the metabolic enzyme monoamine oxidase-A (MAO-A). A common polymorphism in the MAO-A gene, resulting in a relatively low-activity enzyme, has been associated with increased risk for violent or antisocial behavior as well as for depression and anxiety (Caspi et al., 2002; Kim-Cohen et al., 2006).

A recent fMRI study reported that the low-activity MAO-A allele is associated with relatively exaggerated amygdala reactivity and diminished prefrontal regulation of the amygdala (Meyer-Lindenberg et al., 2006). The magnitude of functional coupling between these regions predicted levels of temperamental anxiety, suggesting that the genetic association between the MAO-A low-activity variant and abnormal behavior might be mediated through this circuit. Interestingly, both the 5-HTTLPR S and MAO-A low-activity alleles presumably result in relatively increased 5-HT signaling and exaggerated amygdala reactivity. As the directionality of these effects are consistent with animal studies documenting anxiogenic effects of 5-HT (Burghardt, Bush, McEwen, & LeDoux, 2007; Burghardt, Sullivan, McEwen, Gorman, & LeDoux, 2004; Forster et al., 2006), as well as pharmacological neuroimaging studies demonstrating a potentiation of amygdala reactivity subsequent to acute 5-HT reuptake blockade (Bigos et al., 2008), the imaging genetics data provide important insight regarding the neurobiological and behavioral effects of 5-HT.

Tryptophan Hydroxylase-2 Recent imaging genetics studies that have examined the impact of variation in 5-HT subsystems highlight the potential reciprocal nature by which functional imaging and molecular genetics approaches can be mutually informative in advancing our understanding of the biological mechanism of behavior. Tryptophan hydroxylase-2 (TPH2) is the rate-limiting enzyme in the synthesis of neuronal 5-HT and thus plays a key role in regulating 5-HT neurotransmission (Walther & Bader, 2003; Zhang, Beaulieu, Sotnikova, Gainetdinov, & Caron, 2004). A recent study found that a SNP in the regulatory region of the human TPH2 gene affects amygdala function. Specifically, in comparison with the G allele of this polymorphism, the T allele was associated with relatively exaggerated amygdala reactivity (Brown et al., 2005). This report provides further insight into the biological significance of TPH2 in the human central nervous system and provides a critical next step in our understanding of the importance of this newly identified second tryptophan hydroxylase isoform for human brain function. Moreover, it marks an important advance in the application of functional neuroimaging to the study of genes, brain, and behavior. In contrast to previous studies of genetic effects on brain function, in which the molecular and cellular effects of the candidate variants had been previously demonstrated (e.g., 5-HTTLPR, MAO-A, COMT, and BDNF), these fMRI data provide the first evidence for potential functionality of a novel candidate polymorphism. In this way, the initial identification of a systems-level effect of a specific polymorphism provides impetus for the subsequent characterization of its functional effects at the molecular and cellular level. Converging with this initial imaging genetics finding (and a subsequent replication (Canli, Congdon, Gutknecht, Constable, & Lesch, 2005)), a recent molecular study has demonstrated that this SNP and another promoter SNP impact transcriptional regulation of TPH2 that may affect enzyme availability and 5-HT biosynthesis (Chen, Vallender, & Miller, 2008). Such scientific reciprocity between imaging and molecular genetics illustrates how the contributions of abnormalities in candidate neural systems to complex behaviors and emergent phenomena, possibly including psychiatric illnesses, can be understood from the perspective of their neurobiological origins.

Neuropeptide Y In addition to the above candidate genes, imaging genetics studies have recently explored the impact of genetic variation beyond 5-HT. For example, neuropeptide Y (NPY), a 36-amino-acid peptide neurotransmitter, is an evolutionarily highly conserved molecular component involved in the regulation of brain systems processing stress and emotion (Heilig et al., 1993; Holmes, Heilig, Rupniak, Steckler, & Griebel, 2003). Anxiolytic-like effects of NPY have been reported in a wide range of pharmacologically validated animal models (Broqua, Wettstein, Rocher, Gauthier-Martin, & Junien, 1995; Heilig, Soderpalm, Engel, & Widerlov, 1989; Heilig & Widerlov, 1990; Sajdyk, Vandergriff, & Gehlert, 1999), and NPY release is profoundly induced by stress (Thorsell, Carlsson, Ekman, & Heilig, 1999). In humans, both cerebrospinal fluid and plasma NPY levels correlate with anxiety and stress levels (Boulenger et al., 1996; Irwin et al., 1991; Widerlov, Lindstrom, Wahlestedt, & Ekman, 1988).

A relatively common NPY diplotype consistently predicts NPY mRNA in postmortem brain and lymphoblasts as well as plasma concentrations of NPY (Zhou et al., 2008). Diplotype expression is also inversely proportional to temperamental anxiety (Zhou et al., 2008). Similar to the effect on trait anxiety, NPY haplotype predicts amygdala reactivity in gene-dosage (stepwise) fashion, with heterozygous individuals intermediate in activation (Zhou et al., 2008). Importantly, the magnitude of amygdala activation predicts measures of temperamental anxiety in this sample. Together, the results suggest that NPY effects on temperamental anxiety are mediated in part through biased amygdala reactivity. In addition to these effects, task-related hippocampal activation was predicted by NPY haplotype. This finding is of interest because the functional interactions of the amygdala and hippocampus are critical for emotional memories, and long-lasting changes in hippocampal architecture are induced by stress.

Summary and future directions

Imaging genetics is in its infancy, as the number of genes that have been explored is very few and the strategies for looking at genetically driven variability in brain function are relatively simplistic. Nevertheless, the results above underscore the power and utility of this integrated approach for the identification of biological mechanisms and pathways that mediate individual differences in complex behaviors and related vulnerability to disease. While current imaging genetics studies highlight a powerful new approach to the study of genes, brain, and behavior, the true potential of this approach will be realized only by aggressively expanding the scope and scale of the experimental protocols.

Although single-gene effects on brain function can be readily documented in small samples ($N < 20$), the contributions of multiple genes acting in response to variable environmental pressures is ultimately necessary for the development of truly predictive markers that account for the majority of variance in any given phenotype, such as stress resiliency. For example, the interactive effect of the BDNF Val66Met and 5-HTTLPR on corticolimbic circuitry has been examined recently in an imaging genetics sample of over 100 subjects (Pezawas et al., 2008). An epistatic mechanism between these molecules is suggested by pharmacological and animal models linking 5-HTT and BDNF in cell

signaling related to stress-mediated neuroplasticity (Luellen, Bianco, Schneider, & Andrews, 2006; Ren-Patterson et al., 2006). Surprisingly, the BDNF Met66 allele, which is associated with abnormal regulated BDNF release and reduced hippocampal activity, appears to block the effects of the 5-HTTLPR S allele on reduced amygdala volume. Presumably, the reduced responsivity of the Met66 allele protects against the exaggerated 5-HT signaling associated with the 5-HTTLPR S allele. Such studies provide an example of the biologic epistasis that likely underlies the pathogenesis of a complex disease in the human brain.

Combining existing neuroimaging modalities is another important future direction for imaging genetics. Implementation of multimodal strategies is critical for identifying intermediate mechanisms that mediate the effects of genetic polymorphisms on neural circuit function and related behaviors. The potential of multimodal neuroimaging was recently demonstrated in a study that employed both PET and fMRI to identify the impact of 5-HT_{1A} autoreceptor regulation of 5-HT release on amygdala reactivity (Fisher et al., 2006). In the study, adult volunteers underwent [^{11}C] WAY100635 PET to determine 5-HT_{1A} autoreceptor binding potential, an in vivo index of receptor density. During the same day, all subjects also underwent fMRI to determine the functional reactivity of the amygdala. Remarkably, the density of 5-HT_{1A} autoreceptors accounted for 30–44% of the variability in amygdala reactivity. Downstream effects on 5-HT_{1A} autoreceptors—notably, reduced receptor density—have been hypothesized to mediate neural and behavioral changes associated with the 5-HTTLPR S allele (David et al., 2005). Thus these findings suggest that 5-HT_{1A} autoreceptor regulation of corticolimbic circuitry represents a key molecular mechanism mediating the effects of the 5-HTTLPR.

Ultimately, we anticipate that such mechanistic understanding will allow for the early identification of individuals who are at greater risk for behavioral problems that can have long-term health-related implications. Continued imaging genetics research at the interface of genes, brain, and behavior holds great promise in further explicating the neurobiological mechanisms through which risk for psychiatric disease emerges in the context of environmental adversity (Caspi & Moffitt, 2006; Hariri & Holmes, 2006). Such knowledge will, in turn, facilitate the development of therapeutic interventions, tailored to individual neurobiologies, which will be more effective in combating the enormous personal and public health burden associated with common psychiatric disorders.

ACKNOWLEDGMENTS Portions of this chapter are adapted from Hariri and Weinberger (2003b) and Bigos and Hariri (2007). The author is supported, in part, by a grant from the National Institute of Mental Health Grant (MH072837) and a Young Investigator Award from the National Alliance for Research on Schizophrenia and Depression.

REFERENCES

Bertolino, A., Arciero, G., Rubino, V., Latorre, V., De Candia, M., Mazzola, V., et al. (2005). Variation of human amygdala response during threatening stimuli as a function of 5′HTTLPR genotype and personality style. *Biol. Psychiatry*, *57*(12), 1517–1525.

Bigos, K. L., & Hariri, A. R. (2007). Neuroimaging: Technologies at the interface of genes, brain, and behavior. *Neuroimaging Clin. N. Am.*, *17*(4), 459–467, viii.

Bigos, K. L., Pollock, B. G., Aizenstein, H., Fisher, P. M., Bies, R. R., & Hariri, A. R. (2008). Acute 5-HT reuptake blockade potentiates human amygdala reactivity. *Neuropsychopharmacology*, *33*, 3221–3225.

Boulenger, J. P., Jerabek, I., Jolicoeur, F. B., Lavallee, Y. J., Leduc, R., & Cadieux, A. (1996). Elevated plasma levels of neuropeptide Y in patients with panic disorder. *Am. J. Psychiatry*, *153*(1), 114–116.

Broqua, P., Wettstein, J. G., Rocher, M. N., Gauthier-Martin, B., & Junien, J. L. (1995). Behavioral effects of neuropeptide Y receptor agonists in the elevated plus-maze and fear-potentiated startle procedures. *Behav. Pharmacol.*, *6*(3), 215–222.

Brown, S. M., Manuck, S. B., Flory, J. D., & Hariri, A. R. (2006). Neural basis of individual differences in impulsivity: Contributions of corticolimbic circuits for behavioral arousal and control. *Emotion*, *6*(2), 239–245.

Brown, S. M., Peet, E., Manuck, S. B., Williamson, D. E., Dahl, R. E., Ferrell, R. E., & Hariri, A. R. (2005). A regulatory variant of the human tryptophan hydroxylase-2 gene biases amygdala reactivity. *Mol. Psychiatry*, *10*(9), 884–888.

Burghardt, N. S., Bush, D. E. A., McEwen, B. S., & LeDoux, J. E. (2007). Acute selective serotonin reuptake inhibitors increase conditioned fear expression: Blockade with a 5-HT_{2C} receptor antagonist. *Biol. Psychiatry*, *62*, 1111–1118.

Burghardt, N. S., Sullivan, G. M., McEwen, B. S., Gorman, J. M., & LeDoux, J. E. (2004). The selective serotonin reuptake inhibitor citalopram increases fear after acute treatment but reduces fear with chronic treatment: A comparison with tianeptine. *Biol. Psychiatry*, *55*(12), 1171–1178.

Burmeister, M., McInnis, M. G., & Zollner, S. (2008). Psychiatric genetics: Progress amid controversy. *Nat. Rev. Genet.*, *9*(7), 527–540.

Canli, T., Congdon, E., Gutknecht, L., Constable, R. T., & Lesch, K. P. (2005). Amygdala responsiveness is modulated by tryptophan hydroxylase-2 gene variation. *J. Neural Transm.*, *112*(11), 1479–1485.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., et al. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*(5582), 851–854.

Caspi, A., & Moffitt, T. E. (2006). Gene-environment interactions in psychiatry: Joining forces with neuroscience. *Nat. Rev. Neurosci.*, *7*(7), 583–590.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, *301*(5631), 386–389.

Chen, G. L., Vallender, E. J., & Miller, G. M. (2008). Functional characterization of the human TPH2 5′ regulatory region: Untranslated region and polymorphisms modulate gene expression in vitro. *Hum. Genet.*, *122*(6), 645–657.

David, S. P., Murthy, N. V., Rabiner, E. A., Munafo, M. R., Johnstone, E. C., Jacob, R., et al. (2005). A functional genetic

variation of the serotonin (5-HT) transporter affects 5-HT1A receptor binding in humans. *J. Neurosci.*, *25*(10), 2586–2590.

De Martino, B., Strange, B. A., & Dolan, R. J. (2008). Noradrenergic neuromodulation of human attention for emotional and neutral stimuli. *Psychopharmacology (Berl.)*, *197*(1), 127–136.

Fisher, P. M., Meltzer, C. C., Ziolko, S. K., Price, J. C., & Hariri, A. R. (2006). Capacity for 5-HT_{1A}-mediated autoregulation predicts amygdala reactivity. *Nat. Neurosci.*, *9*(11), 1362–1363.

Forster, G. L., Feng, N., Watt, M. J., Korzan, W. J., Mouw, N. J., Summers, C. H., et al. (2006). Corticotropin-releasing factor in the dorsal raphe elicits temporally distinct serotonergic responses in the limbic system in relation to fear behavior. *Neuroscience*, *141*(2), 1047–1055.

Garpenstrand, H., Annas, P., Ekblom, J., Oreland, L., & Fredrikson, M. (2001). Human fear conditioning is related to dopaminergic and serotonergic biological markers. *Behav. Neurosci.*, *115*(2), 358–364.

Haas, B. W., Omura, K., Constable, R. T., & Canli, T. (2007). Emotional conflict and neuroticism: Personality-dependent activation in the amygdala and subgenual anterior cingulate. *Behav. Neurosci.*, *121*(2), 249–256.

Hariri, A. R., Brown, S. M., Williamson, D. E., Flory, J. D., de Wit, H., & Manuck, S. B. (2006). Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *J. Neurosci.*, *26*(51), 13213–13217.

Hariri, A. R., & Holmes, A. (2006). Genetics of emotional regulation: The role of the serotonin transporter in neural function. *Trends Cogn. Sci.*, *10*(4), 182–191.

Hariri, A. R., Mattay, V. S., Tessitore, A., Fera, F., Smith, W. G., & Weinberger, D. R. (2002). Dextroamphetamine modulates the response of the human amygdala. *Neuropsychopharmacology*, *27*(6), 1036–1040.

Hariri, A. R., Mattay, V. S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., et al. (2002). Serotonin transporter genetic variation and the response of the human amygdala. *Science*, *297*(5580), 400–403.

Hariri, A. R., & Weinberger, D. R. (2003a). Functional neuroimaging of genetic variation in serotonergic neurotransmission. *Genes Brain Behav.*, *2*(6), 314–349.

Hariri, A. R., & Weinberger, D. R. (2003b). Imaging genomics. *Br. Med. Bull.*, *65*, 259–270.

Harmer, C. J., Mackay, C. E., Reid, C. B., Cowen, P. J., & Goodwin, G. M. (2006). Antidepressant drug treatment modifies the neural processing of nonconscious threat cues. *Biol. Psychiatry*, *59*(9), 816–820.

Heilig, M., McLeod, S., Brot, M., Heinrichs, S. C., Menzaghi, F., Koob, G. F., et al. (1993). Anxiolytic-like action of neuropeptide Y: Mediation by Y1 receptors in amygdala, and dissociation from food intake effects. *Neuropsychopharmacology*, *8*(4), 357–363.

Heilig, M., Soderpalm, B., Engel, J. A., & Widerlov, E. (1989). Centrally administered neuropeptide Y (NPY) produces anxiolytic-like effects in animal anxiety models. *Psychopharmacology (Berl.)*, *98*(4), 524–529.

Heilig, M., & Widerlov, E. (1990). Neuropeptide Y: An overview of central distribution, functional aspects, and possible involvement in neuropsychiatric illnesses. *Acta Psychiatr. Scand.*, *82*(2), 95–114.

Holmes, A., Heilig, M., Rupniak, N. M., Steckler, T., & Griebel, G. (2003). Neuropeptide systems as novel therapeutic targets for depression and anxiety disorders. *Trends Pharmacol. Sci.*, *24*(11), 580–588.

Irwin, M., Brown, M., Patterson, T., Hauger, R., Mascovich, A., & Grant, I. (1991). Neuropeptide Y and natural killer cell activity: Findings in depression and Alzheimer caregiver stress. *Faseb J.*, *5*(15), 3100–3107.

Johnstone, T., Somerville, L. H., Alexander, A. L., Oakes, T. R., Davidson, R. J., Kalin, N. H., et al. (2005). Stability of amygdala BOLD response to fearful faces over multiple scan sessions. *Neuroimage*, *25*(4), 1112–1123.

Kendler, K. S., Kuhn, J. W., Vittum, J., Prescott, C. A., & Riley, B. (2005). The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: A replication. *Arch. Gen. Psychiatry*, *62*(5), 529–535.

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch. Gen. Psychiatry*, *60*(9), 929–937.

Kim-Cohen, J., Caspi, A., Taylor, A., Williams, B., Newcombe, R., Craig, I. W., et al. (2006). MAOA, maltreatment, and gene-environment interaction predicting children's mental health: New evidence and a meta-analysis. *Mol. Psychiatry*, *11*(10), 903–913.

Luellen, B. A., Bianco, L. E., Schneider, L. M., & Andrews, A. M. (2006). Reduced brain-derived neurotrophic factor is associated with a loss of serotonergic innervation in the hippocampus of aging mice. *Genes Brain Behav.*, *6*(5), 482–490.

Manuck, S. B., Brown, S. M., Forbes, E. E., & Hariri, A. R. (2007). Temporal stability of individual differences in amygdala reactivity. *Am. J. Psychiatry*, *164*(10), 1613–1614.

Meyer-Lindenberg, A., Buckholtz, J. W., Kolachana, B., Hariri, A. R., Pezawas, L., Blasi, G., et al. (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proc. Natl. Acad. Sci. USA*, *103*(16), 6269–6274.

Munafo, M. R., Brown, S. M., & Hariri, A. R. (2008). Serotonin transporter (5-HTTLPR) genotype and amygdala activation: A meta-analysis. *Biol. Psychiatry*, *63*(9), 852–857.

Munafo, M. R., Clark, T., & Flint, J. (2005). Does measurement instrument moderate the association between the serotonin transporter gene and anxiety-related personality traits? A meta-analysis. *Mol. Psychiatry*, *10*(4), 415–419.

Pezawas, L., Meyer-Lindenberg, A., Drabant, E. M., Verchinski, B. A., Munoz, K. E., Kolachana, B. S., et al. (2005). 5-HTTLPR polymorphism impacts human cingulate-amygdala interactions: A genetic susceptibility mechanism for depression. *Nat. Neurosci.*, *8*(6), 828–834.

Pezawas, L., Meyer-Lindenberg, A., Goldman, A. L., Verchinski, B. A., Chen, G., Kolachana, B. S., et al. (2008). Evidence of biologic epistasis between BDNF and SLC6A4 and implications for depression. *Mol. Psychiatry*, *13*(7), 709–716.

Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science*, *264*(5166), 1733–1739.

Ren-Patterson, R. F., Cochran, L. W., Holmes, A., Lesch, K. P., Lu, B., & Murphy, D. L. (2006). Gender-dependent modulation of brain monoamines and anxiety-like behaviors in mice with genetic serotonin transporter and BDNF deficiencies. *Cell. Mol. Neurobiol.*, *26*(4–6), 753–778.

Rhodes, R. A., Murthy, N. V., Dresner, M. A., Selvaraj, S., Stavrakakis, N., Babar, S., et al. (2007). Human 5-HT transporter availability predicts amygdala reactivity in vivo. *J. Neurosci.*, *27*(34), 9233–9237.

Sajdyk, T. J., Vandergriff, M. G., & Gehlert, D. R. (1999). Amygdalar neuropeptide Y Y1 receptors mediate the anxiolytic-

like actions of neuropeptide Y in the social interaction test. *Eur. J. Pharmacol.*, *368*(2–3), 143–147.

Schinka, J. A., Busch, R. M., & Robichaux-Keene, N. (2004). A meta-analysis of the association between the serotonin transporter gene polymorphism (5-HTTLPR) and trait anxiety. *Mol. Psychiatry*, *9*(2), 197–202.

Sen, S., Burmeister, M., & Ghosh, D. (2004). Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *Am. J. Med. Genet.*, *127B*(1), 85–89.

Somerville, L. H., Kim, H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2004). Human amygdala responses during presentation of happy and neutral faces: Correlations with state anxiety. *Biol. Psychiatry*, *55*(9), 897–903.

Thorsell, A., Carlsson, K., Ekman, R., & Heilig, M. (1999). Behavioral and endocrine adaptation, and up-regulation of NPY expression in rat amygdala following repeated restraint stress. *Neuroreport*, *10*(14), 3003–3007.

Walther, D. J., & Bader, M. (2003). A unique central tryptophan hydroxylase isoform. *Biochem. Pharmacol.*, *66*(9), 1673–1680.

Widerlov, E., Lindstrom, L. H., Wahlestedt, C., & Ekman, R. (1988). Neuropeptide Y and peptide YY as possible cerebrospinal fluid markers for major depression and schizophrenia, respectively. *J. Psychiatr. Res.*, *22*(1), 69–79.

Zhang, X., Beaulieu, J. M., Sotnikova, T. D., Gainetdinov, R. R., & Caron, M. G. (2004). Tryptophan hydroxylase-2 controls brain serotonin synthesis. *Science*, *305*(5681), 217.

Zhou, Z., Zhu, G., Hariri, A. R., Enoch, M. A., Scott, D., Sinha, R., et al. (2008). Genetic variation in human NPY expression affects stress response and emotion. *Nature*, *452*(7190), 997–1001.

65 Components of a Social Brain

JASON P. MITCHELL AND TODD F. HEATHERTON

ABSTRACT Human ecology is radically different from that of other animals, despite the relatively short period of time that separates us phylogenetically from other primates. Recent commentators have suggested that the unique cognitive skills possessed by humans may, in fact, reduce to a small number of primary adaptations for one specialized ability: social cognition, the ability to interact effectively and safely with conspecifics. Emerging research from the neurosciences has begun to elucidate the component parts of these broad social skills. Here, we review evidence that suggests that human social cognition comprises four specialized abilities: a coherent sense of self, the ability to keep track of the mental states of others, control over socially inappropriate emotions and impulses, and sensitivity to threats of exclusion or aggression from other people. We conclude with a review of recent neuroimaging findings that support the view that social cognition has a privileged status in the human cognitive repertoire.

Compared with the ecology of other animals, the lifestyle of *Homo sapiens* is certainly a strange and glamorous affair. Among the world's fauna, humans are the only species that routinely communicates through language, develops reusable tools, designs and constructs architectural structures, tailors its own clothing, engages in economic exchange, composes music, organizes governments, or worships the divine. The sheer range of differences between the species-typical behavior of humans and that of other animals has prompted some theorists to wonder whether sufficient time separates us from our primate cousins for the evolution of distinct cognitive mechanisms to subserve each of these uniquely human pursuits (Tomasello, 1999; Tomasello, Carpenter, Call, Behne, & Moll, 2005). That is, given the relatively small amount of time since the divergence of the human and chimpanzee lineages (between four million and six million years), could separate cognitive adaptations have evolved to support human language and tool use, our insistence on building domestic structures and wearing clothes, and our strange penchant for music, politics, and religion? Or is the divergence between humans and other animals the result of a small number of more primary adaptations, which in turn allowed the development of secondary traits that piggyback on these more fundamental species-unique skills?

JASON P. MITCHELL Department of Psychology, Harvard University, Cambridge, Massachusetts

TODD F. HEATHERTON Department of Psychological and Brain Sciences, Center for Cognitive Neuroscience, Dartmouth College, Hanover, New Hampshire

Surveying the problem posed by the enormous ecological gap between humans and chimpanzees, Tomasello (1999; Tomasello et al., 2005) has argued that many of the unique cognitive skills possessed by humans may, in fact, reduce to a small number of primary adaptations centered on one particularly specialized feature of the mind: the processes that give rise to social cognition. Critical for the development of many of our species-unique behavior—such as language, the use of handheld tools, the assembling of complex social systems—are the ability to learn from and the motivation to teach others. Unlike other primate species, which do not regularly engage in direct instruction, humans actively seek to impart their own knowledge to other humans, especially to young members of the species. Conversely, humans also seem to have a particular motivation to unlock the thoughts and knowledge of others around them, as evidenced by the never-ending "why?" questions asked by two-year-olds. In contrast, whereas apes may acquire new skills from others through imitation, they neither seek out pedagogic contexts nor appear inspired to aid the acquisition of knowledge by others, for example, by explicitly demonstrating a behavior, by repeating it for others to see more clearly, or by taking active steps to shape another's attempts at acquiring a new motor skill.

This species-unique proclivity for learning and teaching imbues humans with enormous behavioral advantages over other animals. Rather than each of us being required to discover facts about the world for ourselves, humans routinely capitalize on the accumulated knowledge of all the other people to whom they have some access. That is, whereas other apes can learn about only those objects and events with which they come into immediate contact, humans can avail themselves of what others, both past and present, have uncovered about the world. As such, we tend to add complexity to already existing knowledge rather than generating it from scratch—such as improving on the workings of already-existing tools and behaviors instead of creating them de novo. Unlike the knowledge of apes, which effectively remains trapped inside the mind of its possessor, human know-how spreads quickly through any population of humans who are in contact with each other, so if one person acquires a new bit of knowledge, all other humans in her group can generally make use of it as well.

What grants humans the ability to traffic so readily in the knowledge and beliefs of others? That is, starting with

the basic primate mental algorithm, what additional cognitive software is needed to create a social animal of the human variety? Here, we expand on a model that attempts to identify the necessary basic components of the social brain (Heatherton, in press; Krendl & Heatherton, in press). From this perspective, building a social brain requires at least four distinct cognitive abilities. The first of these is a stable and coherent sense of self, the knowledge that one has mental states that are both idiosyncratic and private but that are in principle capable of being shared with others. The motivation to engage in teaching behavior is predicated on the assumption that one's knowledge is idiosyncratic and not immediately apparent to others, that is, that one has knowledge that is unique to a self. In addition, teachers must assume that their learners are sufficiently similar to them that the teachers can impart their knowledge, that is, that learners are capable of thinking the same thoughts as oneself; the reason we do not routinely teach calculus to first-graders or lecture our pets about quantum physics is that we do not assume that their minds are sufficiently similar to our own to allow them to think the thoughts that we do.

Second, if one adaptive goal of the social brain is to foster the spread of knowledge from one mind to the next, humans require a mechanism for keeping track of the mental states of others. This skill, variously known as *mentalizing* or *theory of mind*, allows one person to intuit the beliefs, thoughts, feelings, goals, and desires of other people for the purposes of both predicting and influencing their behavior.

Third, human knowledge spreads only among individuals who are in contact with one another. As such, one requirement for human social life is the ability to live in close proximity with large numbers of other people. Although other primates may form groups as large as a few hundred individuals (Marc Hauser, personal communication, May 21, 2008), humans have developed the ability to live in close contact with a number of conspecifics that exceeds this by several orders of magnitude (witness New York City and Tokyo as examples). Living in large social groups, even relatively small human groups on the order of a rural village, poses cognitive challenges of considerable difficulty. Among the most substantial of these is the need to regulate one's own desires and emotions to maintain social harmony. That is, maintaining cooperative relationships among large numbers of unrelated individuals requires a cognitive system with special abilities to inhibit potentially destructive emotional expression, a skill with which other primates (with the possible exception of bonobos) struggle.

Finally, the centrality of group living to human ecology generates two sources of special dangers that the social brain must have evolved to handle: threats from one's own group and threats from other groups. Threats from one's own group generally take the form of the potential for social exclusion, a fate that at many times in human history was equivalent to a death sentence. Conversely, because humans naturally form groups, one always faces the possibility of intergroup competition for resources or explicit conflict (group raids, war, etc.). Accordingly, the social brain should be expected to have developed mechanisms for anticipating and dealing with both sources of intraspecies threat.

In this chapter, we explore the functional neuroanatomy associated with these four components of social cognition brain: self-awareness, theory of mind, self-regulation, and threat detection. Unlike many other aspects of cognition, almost everything we know about the social brain has been uncovered in the last decade and a half. Fortunately, the emergence of social neuroscience has been both rapid and far-reaching; therefore, despite its infancy, this approach has netted a substantial number of reliable and surprising empirical findings about how the brain gives rise to human sociality. To this end, we conclude the chapter with a brief speculative review of some unexpected observations of the peculiar "specialness" of social thought within the human cognitive repertoire.

Components of the social brain

Self-Awareness Survival in human social groups requires that people monitor their behavior and thoughts and evaluate them against prevailing group (social) norms. Discussions of the importance of such introspective awareness have a long history in psychology. For example, William James (1890) devoted a chapter of his *Principles of Psychology* to issues of one's own knowledge of self. More recently, Neisser (1988) has described a number of distinct ways in which one can think of selfhood, such as a "conceptual self" that represents our understanding of our own personality traits and dispositions, an "ecological self" that represents our sense of authoring our own actions in the environment, and a "narrative self" that maintains our sense of personal history and autobiographical memory (also see Boyer, Robbins, & Jack, 2005; S. Gallagher, 2000). Sedikides and Skowronski (1997) argue that this last sense of self is a widely shared human trait that leads to more efficient mental processing of personal and contextual information, thereby increasing the likelihood of survival and reproduction.

Cognitive neuroscience has made considerable progress in identifying the functional neuroanatomy underlying several of these different aspects of self-awareness. Of these, most is known about the brain regions subserving introspective knowledge of one's own stable personality traits and dispositions. Specifically, both neuroimaging research and studies with neuropsychological patients have implicated ventral aspects of the medial prefrontal cortex (vMPFC) as contributing importantly to conceptual aspects of selfhood. For example, a considerable number of neuroimaging studies

have implicated this region in tasks that require participants to judge their own personality traits (Craik et al., 1999; Heatherton et al., 2006; Johnson et al., 2002; Kelley et al., 2002; Macrae, Moran, Heatherton, Banfield, & Kelley, 2004; Moran, Macrae, Heatherton, Wyland, & Kelley, 2006; Ochsner et al., 2004; Schmitz, Kawahara-Baccus, & Johnson, 2004; Zysset, Huber, Ferstl, & von Cramon, 2002) or report on their preferences and opinions (Ames, Jenkins, Banaji, & Mitchell, 2008; Jenkins, Macrae, & Mitchell, 2008; Mitchell, Macrae, & Banaji, 2006), compared to judging these characteristics in others.

The more an item is believed to be reflective of the self, the greater the activity in this area (Moran et al., 2006), and items that are inherently self-relevant (such as personal information) lead to vMPFC activity even during passive viewing when people are not asked to engage in self-referential processing (Moran, Heatherton, & Kelley, in press). Interestingly, damage to this region can lead to deficits in the organization of knowledge about one's preferences. Fellows and Farah (2007) have reported that, when asked to indicate how much they like/dislike various stimuli, patients with vMPFC lesions show unusually large discrepancies between testing sessions, suggesting that damage to this region leads either to failure to retrieve knowledge of one's preferences or to instability in otherwise stable aspects of selfhood.

In addition to knowledge about one's own personality and preferences, humans require a system for keeping track of the actions they perform (what Neisser referred to as an "ecological self"). In recent behavioral work, researchers have demonstrated enhanced memory performance for stimuli that a participant freely selects, compared to those that are selected by someone else (Cloutier & Macrae, 2007). Extending these results, Powell, Macrae, Cloutier, Metcalfe, and Mitchell (submitted) demonstrated that the sense of agentic free selection was associated with greater activity in the intraparietal sulcus (IPS), a region that was previously implicated in maintaining representations of one's goals and computing the motor commands needed to bring them about (Grafton & Hamilton, 2007; Hamilton & Grafton, 2006; Tunik, Rice, Hamilton, & Grafton, 2007). Moreover, activity in IPS correlated with the mnemonic fate of items—showing greater activity for items that were later remembered than for those that were later forgotten—but only for items that were freely chosen by the participant. In other words, much as the conceptual sense of who one is relative to others—that is, one's stable traits and preferences—relies on vMPFC, the agentic sense of what one has done relative to others draws on the activity of IPS.

Theorists have also pointed out the importance to the self of feeling continuity between one's present and past experiences, that is, of possessing a consistent autobiography. Studies of autobiographical memory have implicated a wide range of brain areas, including retrosplenial cortex, parahippocampal gyrus, temporoparietal junction (TPJ), medial frontal cortex, temporal pole, cerebellum, and the hippocampus (for review, see Maguire, 2001). Of these, perhaps the most consistent finding is that of the MPFC, which has regularly been observed during tasks in which participants are asked to remember events from their own personal past.

Although there exists no specific "self" spot in the brain, this brief review of the cognitive neuroscience of self-awareness suggests that specific regions, such as MPFC and IPS, play an important role in distinguishing between self and other, a critical feature of the adaptive social brain. An important current goal in social neuroscience is further description of the precise cognitive processes that are subserved by these regions. For example, how exactly does the MPFC compute one's preferences or retrieve knowledge about one's stable personality traits? How does the IPS "tag" actions as having been performed by self rather than by another person? Elucidating the computations carried out by these regions and how they combine to give rise to our sense of self promises significant advances to our understanding of how the self contributes to social cognition.

Mentalizing One of the most critical components of social cognition is the ability to infer the mental states of other people, a skill that is alternately referred to as *mentalizing*, *having a theory of mind*, or *adopting the intentional stance* (Dennett, 1987). Mentalizing enables the ability to empathize and cooperate with others, accurately interpret other people's behavior, and even deceive others when necessary. A rapidly emerging neuroimaging literature on mentalizing has consistently implicated a small number of regions in making inferences about the mental characteristics of other people: MPFC, (TPJ), and medial parietal cortex (Amodio & Frith, 2006; Frith & Frith, 1999, 2001; H. L. Gallagher & Frith, 2003; Mitchell, 2006).

But how exactly does one make sense of the thoughts and feelings of other people, given that perceivers never have direct access to the inner workings of another person? Some cognitive scientists and philosophers have argued that one solution to this problem of mentalizing is suggested by the fact that, although one cannot directly perceive the mental states of another person, one does typically have immediate access to a decent proxy system: oneself. Simulationist (or projectionist) accounts of mentalizing posit that one way in which perceivers may infer others' goals, feelings, or preferences is to put themselves in the same situation as the target person, read off the feelings and thoughts that accompany that simulation, and then attribute (roughly) the same mental states to the target individual. However, this strategy for using oneself as a proxy for others works only when one can reasonably assume that another person will have similar responses to a situation; if a perceiver believes himself or

herself to be highly dissimilar from a target individual, the use of self-referential mentalizing can be inappropriate.

Interestingly, a series of neuroimaging studies has provided evidence in favor of this view by capitalizing on the well-characterized role of vMPFC in self-referential thought, as reviewed above (Jenkins et al., 2008; Mitchell, Banaji, & Macrae, 2005; Mitchell et al., 2006). Specifically, when perceivers mentalized about the preferences and opinions of a similar other (e.g., someone who shared the same social and political attitudes), a region of vMPFC was engaged; the same region was highly engaged when perceivers considered their own preferences. In contrast, a more dorsal region of MPFC was preferentially engaged when perceivers mentalized about dissimilar others. These results suggest that perceivers might indeed draw on their own knowledge about self in understanding the mental states of others, at least those others who are believed to be relevantly similar.

How does one go about inferring the mental states of those perceived to be dissimilar from self, that is, when simulation may be inappropriate? Although the exact mechanisms that support nonsimulationist mentalizing have yet to be detailed, the TPJ appears to augment the role of the MPFC in social cognition and may play an important role in non-self-referential theory of mind. Specifically, the TPJ is preferentially engaged during inferences about particular kinds of mental states, namely, others' beliefs or knowledge about the world (Saxe, Carey, & Kanwisher, 2003; Saxe & Kanwisher, 2003; Saxe & Wexler, 2005), which may be inferred without reference to one's own beliefs or knowledge. For example, many studies of inferences about others' knowledge have used the classic false belief task, in which a character in a story believes that X is true, although the perceiver knows that X is false (e.g., the location of a hidden object). Here, perceivers cannot use their own self-referential knowledge (X is false) to make inferences about the target's beliefs, and the TPJ may subserve some of the ancillary social-cognitive processes that allow this decoupling of one's own knowledge of self from knowledge of others, perhaps by using a more rule-based approach to mentalizing.

Self-Regulation People who defy group norms—such as by cheating, lying, or being incompetent—often experience social emotions that indicate that something is wrong. We feel embarrassed when we goof, guilty when we harm, and ashamed when we get caught. Such social emotions serve as important guides for subsequent behavior (Baumeister, Stillwell, & Heatherton, 1994); for example, feeling embarrassed or ashamed motivates behavior to repair social relationships, and feeling guilty about considering cheating on one's partner helps to rein in one's response to temptation. In other words, social emotions promote self-regulation, which allows us to alter our behavior, make plans, choose from alternatives, focus attention on pursuit of goals, inhibit competing thoughts, and regulate social behavior (Baumeister, Heatherton, & Tice, 1994). Self-regulation refers not only to executive processes such as working memory, attention, memory, choice, and decision making, but also to the control of emotion (covering issues of affect, drive, and motivation). Although humans have the capacity to delay gratification, control appetites and impulses, and persevere to attain goals, failures of self-regulation are among the most important and perplexing problems facing society (e.g., drug abuse, domestic violence, binge eating).

The importance of self-regulation is that it helps people to control their behaviors and actions so that they remain in good standing within their groups. Throughout evolutionary history, people have faced the continuing struggle between satisfying personal desires and being a good member of the group. From a selfish hedonistic perspective, people should eat as much as they want, freeload on group resources, ignore restrictions on sexual conduct, use substances that induce euphoria, and so on. Essentially, strictly on the grounds of individual enjoyment, people should engage in activities that engage the mesolimbic dopamine system and produce rich feelings of reward. If it feels good, do it. But those who eat all the food, fail to be productive because they are incapacitated, or poach mates are bad group members; therefore most groups have shared norms and standards of conduct that discourage or place constraints on selfish, hedonistic activities. Religious doctrine, common to most cultural groups, places explicit rules on such behavior. Subcortical reward motives, then, are in constant battle with higher-level cognitive beliefs and values.

Neuroscience research indicates that various regions of PFC are responsible for the human capacity for self-regulation (see the review by Banfield, Wyland, Macrae, Munte, & Heatherton, 2004). For instance, functional neuroimaging studies have implicated the ACC in decision monitoring, initiating the selection of an appropriate novel response from several alternatives, performance monitoring, action monitoring, detection or processing of response conflict, and internal cognitive control (Wyland, Kelley, Macrae, Gordon, & Heatherton, 2003). More recently, we found an important role for the ACC in efforts to suppress unwanted thoughts (Mitchell et al., 2007), such that ACC was transiently engaged following the occurrence of unwanted thoughts, whereas dorsolateral PFC was most active during tonic efforts to suppress those thoughts. This finding is in keeping with the important role of prefrontal regions in executive functions more generally, all of which are necessary for successful self-regulation (Miller & Cohen, 2001). Since the case of Phineas Gage (Damasio, Grabowski, Frank, Galaburda, & Damasio, 1994; Macmillan, 2002), we have known that damage to certain prefrontal regions is associated with a lack of impulse control and self-regulatory difficulties more generally. The role of lateral PFC regions in

regulating social emotions appears to be among the most robust findings in social neuroscience.

How is brain activity related to self-regulatory behavior? Failures of self-regulation are commonplace, as can be confirmed by asking chronic dieters who are usually on a diet but seem to sabotage their diets with occasional, or not so occasional, bouts of overeating. In a series of classic studies, Herman and Polivy (1975) demonstrated that chronic dieters are prone to excessive eating in certain situations. In one of their first studies, chronic dieters (called *restrained eaters*) and nondieters were invited to the laboratory to engage in a supposed taste test, which was described as a test of perception (because the researchers did not want the subjects to know that eating was being monitored). Prior to the taste test, some of the participants were asked to drink one or two flavorful but obviously fattening milkshakes. The participants then were asked to taste and rate flavors of ice cream and were invited to help themselves to as much of the ice cream as they wanted. Nondieters ate sensibly; those who did not receive a preload ate a lot of the good-tasting (and free) ice cream, whereas those who had drunk one or two milkshakes ate much less of the ice cream. Restrained eaters, however, did just the opposite, eating much more if they had received a preload. This has been called the "what the hell effect," the mindset of the dieter being "I've blown my diet, so I might as well just keep eating." This finding of disinhibited eating has since been confirmed repeatedly (Heatherton & Baumeister, 1991) and serves as an excellent example of self-regulatory collapse.

In a recent imaging study, Demos, Kelley, and Heatherton (submitted) told dieters ($N = 50$) and nondieters ($N = 50$) that they were investigating the effect of nasal cavity temperature on BOLD signal artifact in orbitofrontal cortex. This cover story provided a rationale for requiring participants to drink either a large glass of cold water or a large glass of a high-calorie chocolate milkshake that would be forbidden in most diets and that had been associated with disinhibited eating by dieters. During scanning, subjects viewed images of animals, environmental scenes, people, and attractive food and made simple person perception judgments (i.e., "Is there a person present in the image?"). The person perception judgment was included to ensure that subjects attended to all images and to further disguise the primary goal of the study in examining cue reactivity to food images. A voxel-wise whole-brain ANOVA revealed that bilateral regions of the nucleus accumbens (NAcc) demonstrated a dietary status × preload interaction that mimicked the well-established behavioral pattern of disinhibited eating. Specifically, when participants received water prior to scanning, nondieters showed greater NAcc activity to food images than did dieters. This pattern was reversed in individuals who received the milkshake preload, such that dieters produced much greater NAcc activity than nondieters did. Activity in the left amygdala revealed the reverse interaction: activity in the left amygdala was greater for dieters than for nondieters following the water and was greater for nondieters than for dieters following the milkshake preload. Importantly, the interactions in NAcc and amygdala were unique to food images; these patterns were not present when subjects viewed any of the nonfood images. This study demonstrates two important points. First, somehow dieters are able to view attractive food cues without activating reward circuitry while their diets are intact (although how they do this is currently unknown). In sharp contrast, following a large milkshake that should have induced satiety (and that eliminated a reward response among nondieters), dieters showed much greater reward-related food cue reactivity.

Studies such as these described begin to provide information relevant to people's efforts to regulate their thoughts and actions. Coupled with research identifying brain regions involved in emotional regulation (Ochsner et al., 2004), social neuroscientists are tackling longstanding questions regarding the human capacity to control itself in order to obtain long term goals and be effective group members.

Detection of Threat

Over the course of human evolution, a major adaptive challenge to survival was other people, both ingroup members and members of other groups. However, the nature of these threats is distinctly different. As was discussed above, the fundamental human need to belong makes social exclusion a potentially fatal sentence. In contrast, members of other groups pose a threat because they represent physical danger or competition for limited resources. As such, the social brain requires mechanisms not only to detect threats posed by both ingroup and outgroup members, but also to differentiate between the specific nature of each kind of social threat. A variety of brain regions have been identified as relevant to the detection of threat, including the amygdala and the anterior cingulate cortex (ACC), both of which have also been implicated in specifically social aspects of threat detection.

If humans have a fundamental need to belong, our system for social cognition must necessarily include mechanisms for detecting inclusionary status (Leary, Tambor, Terdal, & Downs, 1995; Macdonald & Leary, 2005). That is, given the dangers posed by exclusion from one's group, humans must be capable of benchmarking their interpersonal relationships with others. A recent set of neuroimaging studies has examined the concomitants of a particular form of ingroup threat detection: social rejection. In the first of these studies, Eisenberger, Lieberman, and Williams (2003) engineered an experimental situation in which participants were unexpectedly excluded from a computer game by a virtual interaction partner, who simply started ignoring the participant. Participants reported experiencing social rejection under these circumstances, and the authors found that the depth of such

rejection feelings was positively correlated with activity in the dorsal ACC. Since this initial study, other studies have also implicated the ACC, although some of them find a more ventral rather than dorsal region (Somerville, Heatherton, & Kelley, 2006). Another recent study (Burklund, Eisenberger, & Lieberman, 2007) found a relationship between rejection sensitivity and activity in both dorsal ACC and the ventral ACC during emotional processing. The somewhat disparate findings of these studies indicate the need for further research to more clearly identify the neural correlates of states of social distress, especially in terms of the functional roles of ACC in processing and responding to threat cues.

Of course, social threats also come from outside one's own group. The cognitive neuroscience of such external threats has burgeoned in recent years, the amygdala being the area most commonly identified as relevant to outgroup threat. For example, studies have associated amygdala activity with white perceivers' negative responses to African-Americans (Cunningham et al., 2004; Phelps et al., 2000; Richeson et al., 2003). People who possess stigmatizing conditions that make them seem less than human, such as the homeless, also activate regions of the amygdala (Harris & Fiske, 2006). Krendl, Macrae, Kelley, Fugelsang, and Heatherton (2006) have found amygdala responses to physically unattractive individuals or people who are otherwise stigmatized by their appearance.

Considered together, these data strongly suggest that evaluations of outgroup members engage the amygdala. But what precisely does the amygdala do in the context of social cognition? One possibility derives from the longstanding notion that the amygdala may play a special role in responding to stimuli that elicit fear (Blanchard & Blanchard, 1972; Feldman-Barrett & Wager, 2006; LeDoux, 1996), suggesting that the amygdala may contribute to hard-wired circuits that have developed over the course of evolution to protect animals from danger. For example, the amygdala is robustly activated in response to primary biologically relevant stimuli (e.g., faces, odors, tastes), even when these stimuli remain below the subjects' level of reported awareness (Morris, Ohman, & Dolan, 1998; Whalen et al., 1998).

However, many recent imaging studies have also observed amygdala activity in response to stimuli of positive valence, indicating that the amygdala is not solely concerned with fear. Indeed, some have argued that the amygdala is important for drawing attention to novel stimuli that have biological relevance. For instance, Stephan Hamann and colleagues (Hamann, Herman, Nolan, & Wallen, 2004) found that activity within the amygdala increased when both men and women viewed sexually arousing stimuli, such as short film clips of sexual activity or pictures of opposite-sex nudes. As such, the amygdala may play a role in processing social emotions more generally, because such affective states have direct relevance in maintaining long-term social relationships. Along these lines, Paul Whalen has argued that the amygdala is especially concerned with ambiguous stimuli that provide insufficient information to discern the nature of the threat (Whalen, 1998, 2007); since fear faces indicate an unspecified threat, they may engage the amygdala more than less ambiguous facial displays, such as anger (Whalen et al., 2001).

The "special" nature of social cognition

As was discussed above, a particular set of social-cognitive processes supporting our understanding of other minds has consistently been linked to a fairly small number of brain regions: specifically, the MPFC, TPJ, and medial parietal cortex. Interestingly, this set of brain regions is marked by an unusually high rate of resting metabolic activity. That is, different brain areas have overall higher or lower rates of metabolism when individuals rest passively without performing a specified task (Gusnard & Raichle, 2001), suggesting that some regions may chronically carry out cognitive processing even in the absence of explicit task demands. This observation suggests that, when allowed to relax to baseline, the human brain seems to persist in some kinds of social-cognitive processing, subserved by regions such as the MPFC and TPJ that have particularly high resting metabolic rates.

Perhaps even more speculatively, these brain regions demonstrate an unusual tendency to "deactivate" when perceivers engage in tasks that do not rely on social thought. When performing tasks that instead require consideration of nonsocial aspects of the world (e.g., visual search or semantic decision tasks), the human brain seems to suspend the otherwise high resting activity in the MPFC, TPJ, and medial parietal cortex. Suggestively, the tendency to deactivate during nonsocial tasks suggests that the processing in these regions may be in some way incompatible with nonsocial cognition; the brain regions that subserve other aspects of cognition (e.g., language or vision) do not share this feature of active inhibition when one performs a task that does not require those processes. Together, these observations suggest, albeit speculatively, that the human cognitive system may be in a state of continuous readiness to encounter other minds (hence the high resting metabolic rate evinced by these regions) and that this social default must be actively suspended to engage appropriately with nonsocial entities, such as inanimate objects.

Conclusion

Over the past two decades, the integration of cognitive neuroscience and social psychology has led to a spate of insights into the neural basis of human social cognition. In beginning to examine the neural underpinnings of social behavior,

researchers have begun the process of carving social cognition "at its joints," using the brain to identify the cognitive processes that allow humans to tap into the minds of others. We have learned that social thought relies on a suite of different mechanisms, some of which are quite distinct from the processes that subserve nonsocial aspects of human cognitive abilities, whereas others make use of more general-purpose computations. Here, we have specifically suggested that the social brain comprises at least four distinct mechanisms: awareness of self, understanding other minds, regulation of one's own behavior and emotions, and the detection and avoidance of social threat. But the field of social neuroscience is still very much in its infancy, and much remains to be added to this list. We look forward to such future insights with excitement.

ACKNOWLEDGMENTS This work was supported by the following grants: NSF BCS 0642448 to JPM and NIMH 59282, NIDA 022582, and NSF BCS 0354400 to TFH. We acknowledge our collaborators at the Dartmouth Center for Social Brain Sciences and the Harvard University Department of Psychology as collaborators on much of the research described.

REFERENCES

Ames, D. L., Jenkins, A. C., Banaji, M. R., & Mitchell, J. P. (2008). Taking another's perspective increases self-referential neural processing. *Psychol. Sci*, *19*, 642–644.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nat. Rev. Neurosci.*, *7*(4), 268–277.

Banfield, J. F., Wyland, C. L., Macrae, C. N., Munte, T. F., & Heatherton, T. F. (2004). The cognitive neuroscience of self-regulation. In *Handbook of self-regulation: Research, theory, and applications* (pp. 63–83). New York: Guilford Press.

Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). *Losing control: How and why people fail at self-regulation.* San Diego: Academic Press.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychol. Bull.*, *115*(2), 243–267.

Blanchard, D. C., & Blanchard, R. J. (1972). Innate and conditioned reactions to threat in rats with amygdaloid lesions. *J. Comp. Physiol. Psychol.*, *81*(2), 281–290.

Boyer, P., Robbins, P., & Jack, A. I. (2005). Varieties of self-systems worth having. *Conscious. Cogn.*, *14*(4), 647–660.

Burklund, L. J., Eisenberger, N. I., & Lieberman, M. D. (2007). The face of rejection: Rejection sensitivity moderates dorsal anterior cingulate activity to disapproving facial expressions. *Soc. Neurosci.*, *2*(3–4), 238–253.

Cloutier, J., & Macrae, C. N. (2007). The feeling of choosing: Self-involvement and the cognitive status of things past. *Conscious. Cogn.*, doi:10.1016/j.concog.2007.05.010.

Craik, F. I. M., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., et al. (1999). In search of the self: A positron emission tomography study. *Psychol. Sci.*, *10*, 26–34.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Chris Gatenby, J., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychol. Sci.*, *15*(12), 806–813.

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: Clues about the brain from the skull of a famous patient. *Science*, *264*(5162), 1102–1105.

Demos, K. E., Kelley, W. M., & Heatherton, T. F. (submitted). The neural correlates of dietary restraint violations.

Dennett, D. C. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An FMRI study of social exclusion. *Science*, *302*(5643), 290–292.

Feldman-Barrett, L., & Wager, T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. *Curr. Dir. Psychol. Sci.*, *15*, 79–83.

Fellows, L. K., & Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: Judgment under uncertainty or judgment per se? *Cereb. Cortex*, *17*(11), 2669–2674.

Frith, C. D., & Frith, U. (1999). Interacting minds: A biological basis. *Science*, *286*, 1692–1695.

Frith, C. D., & Frith, U. (2001). The biological basis of social interaction. *Curr. Dir. Psychol. Sci.*, *10*, 151–155.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind." *Trends Cogn. Sci.*, *7*(2), 77–83.

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends Cogn. Sci.*, *4*(1), 14–21.

Grafton, S. T., & Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Hum. Movement Sci.*, *26*, 590–616.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nat. Rev. Neurosci.*, *2*, 685–694.

Hamann, S., Herman, R. A., Nolan, C. L., & Wallen, K. (2004). Men and women differ in amygdala response to visual sexual stimuli. *Nat. Neurosci.*, *7*(4), 411–416.

Hamilton, A. F., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *J. Neurosci.*, *26*(4), 1133–1137.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychol. Sci.*, *17*(10), 847–853.

Heatherton, T. F. (in press). Building a social brain. In A. Todorov, S. T. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind.* Oxford, UK: Oxford University Press.

Heatherton, T. F., & Baumeister, R. F. (1991). Binge eating as escape from self-awareness. *Psychol. Bull.*, *110*(1), 86–108.

Heatherton, T. F., Wyland, C. L., Macrae, C. N., Demos, K. E., Denny, B. T., & Kelley, W. M. (2006). Medial prefrontal activity differentiates self from close others. *Soc. Cogn. Affect. Neurosci.*, *1*, 18–25.

Herman, C. P., & Polivy, J. (1975). Anxiety, restraint and eating behavior. *J. Abnorm. Psychol.*, *84*, 666–672.

James, W. (1890). *Principles of psychology.* New York: Holt, Rinehart & Wilson.

Jenkins, A. C., Macrae, C. N., & Mitchell, J. P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proc. Natl. Acad. Sci. USA*, *105*(11), 4507–4512.

Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., & Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain*, *125*(8), 1808–1814.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *J. Cogn. Neurosci.*, *14*(5), 785–794.

Krendl, A. C., & Heatherton, T. F. (in press). Self versus others/ self-regulation. In G. G. Berntson & J. T. Cacioppo (Eds.), *Handbook of neuroscience for the behavioral sciences*. Hoboken, NJ: John Wiley.

Krendl, A. C., Macrae, C. N., Kelley, W. M., Fugelsang, J. F., & Heatherton, T. F. (2006). The good, the bad, and the ugly: An fMRI investigation of the functional anatomic correlates of stigma. *Soc. Neurosci.*, *1*(1), 5–15.

Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *J. Person. Soc. Psychol.*, *68*(3), 518–530.

LeDoux, J. E. (1996). *The emotional brain*. New York: Simon & Schuster.

Macdonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychol. Bull.*, *131*(2), 202–223.

Macmillan, M. (2002). *An odd kind of fame: Stories of Phineas Gage*. Cambridge, MA: MIT Press.

Macrae, C. N., Moran, J. M., Heatherton, T. F., Banfield, J. F., & Kelley, W. M. (2004). Medial prefrontal activity predicts memory for self. *Cereb. Cortex*, *14*(6), 647–654.

Maguire, E. A. (2001). Neuroimaging studies of autobiographical event memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *356*(1413), 1441–1451.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.*, *24*, 167–202.

Mitchell, J. P. (2006). Mentalizing and Marr: An information processing approach to the study of social cognition. *Brain Res.*, *1079*(1), 66–75.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *J. Cogn. Neurosci.*, *17*(8), 1306–1315.

Mitchell, J. P., Heatherton, T. F., Kelley, W. M., Wyland, C. L., Wegner, D. M., & Macrae, C. N. (2007). Separating sustained from transient aspects of cognitive control during thought suppression. *Psychol. Sci.*, *18*(4), 292–297.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*(4), 655–663.

Moran, J. M., Heatherton, T. F., & Kelley, W. M. (in press). Modulation of cortical midline structures by implicit and explicit self-reference evaluation. *Soc. Neurosci.*

Moran, J. M., Macrae, C. N., Heatherton, T. F., Wyland, C. L., & Kelley, W. M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *J. Cogn. Neurosci.*, *18*(9), 1586–1594.

Morris, J. S., Ohman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, *393*(6684), 467–470.

Neisser, U. (1988). Five kinds of self-knowledge. *Philos. Psychol.*, *1*(1), 35–59.

Ochsner, K. N., Knierim, K., Ludlow, D. H., Hanelin, J., Ramachandran, T., Glover, G., et al. (2004). Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *J. Cogn. Neurosci.*, *16*(10), 1746–1772.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.*, *12*(5), 729–738.

Powell, L. J., Macrae, C. N., Cloutier, J., Metcalfe, J., & Mitchell, J. P. (submitted). Dissociable neural substrates for agentic versus conceptual representations of self.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., et al. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nat. Neurosci.*, *6*(12), 1323–1328.

Saxe, R., Carey, S., & Kanwisher, N. (2003). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annu. Rev. Psychol.*, *55*(4), 1–38.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: fMRI investigations of theory of mind. *NeuroImage*, *19*, 1835–1842.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*(10), 1391–1399.

Schmitz, T. W., Kawahara-Baccus, T. N., & Johnson, S. C. (2004). Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *NeuroImage*, *22*(2), 941–947.

Sedikides, C., & Skowronski, J. J. (1997). The symbolic self in evolutionary context. *Pers. Soc. Psychol. Rev.*, *1*(1), 80–102.

Somerville, L. H., Heatherton, T. F., & Kelley, W. M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nat. Neurosci.*, *9*(8), 1007–1008.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behav. Brain Sci.*, *28*, 675–735.

Tunik, E., Rice, N. J., Hamilton, A., & Grafton, S. T. (2007). Beyond grasping: Representation of action in human anterior intraparietal sulcus. *NeuroImage*, *36*(Suppl 2), T77–T86.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Curr. Dir. Psychol. Sci.*, *7*(6), 177–188.

Whalen, P. J. (2007). The uncertainty of it all. *Trends Cogn. Sci.*, *11*, 499–500.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J. Neurosci.*, *18*(1), 411–418.

Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., Wright, C. I., & Rauch, S. L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion*, *1*(1), 70–83.

Wyland, C. L., Kelley, W. M., Macrae, C. N., Gordon, H. L., & Heatherton, T. F. (2003). Neural correlates of thought suppression. *Neuropsychologia*, *41*(14), 1863–1867.

Zysset, S., Huber, O., Ferstl, E., & von Cramon, D. Y. (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *NeuroImage*, *15*(4), 983–991.

66 The Neural Basis of Emotion Regulation: Making Emotion Work for You and Not Against You

JENNIFER S. BEER

ABSTRACT Too much emotion can get you into trouble. However, the maxim "Everything in moderation" applies to our emotional lives because a life without any emotion is not desirable either. We certainly need to experience some emotions in order to recognize when good *and* bad things happen. Luckily, the body's campaign for homeostasis extends to emotional processes; therefore we do have some mechanisms in place to regulate our emotions so that we can maximize their benefit and minimize their costs. What neural circuitry is involved in the control of emotion? Does it depend on the psychological manner in which we strive to regulate our emotions? Is it different from the neural basis of regulatory processes used to control nonemotional processes? This chapter addresses these questions by reviewing the human neuroscience research on three broad categories of emotional regulation: processes of distraction, reappraisal, and controlling emotional influences on decision making.

Too much emotion can get you into trouble. There are certainly extreme examples such as "crimes of passion," in which people may commit acts of violence that would be inconceivable to them in a nonemotional state, or the pursuit of drug-induced euphoria, which cannibalizes resources needed to maintain employment and relationships. In addition to problems of intensity, emotions can often butt heads with societal norms. For example, the gloomy face of the groom's mother at a wedding might reflect her internal emotional state but bode poorly for her relationship with the future daughter-in-law. However, the maxim "Everything in moderation" applies to our emotional lives because a life without any emotion is not desirable either. We certainly need to experience some emotions in order to recognize when good *and* bad things happen. In fact, Eric Wilson argues in his book *Against Happiness: In Praise of Melancholy* that American culture has placed too much emphasis on suppressing negative emotions and boosting positive emotions, which is costing people the motivational and creative lessons that can be learned only from negative experience. Luckily, the body's campaign for homeostasis extends to emotional processes; therefore we do have some mechanisms in place to regulate our emotions so that we can maximize their benefit and minimize their costs. What neural circuitry is involved in the control of emotion? Does it depend on the psychological manner in which we strive to regulate our emotions? Is it different from the neural basis of regulatory processes used to control nonemotional processes? This chapter addresses these questions by reviewing the human neuroscience research on three broad categories of emotional regulation: processes of distraction, reappraisal, and controlling emotional influences on decision making.

JENNIFER S. BEER University of Texas at Austin, Austin, Texas

Definitions and methodological issues for the study of emotion regulation

Emotion regulation is a diverse set of processes aimed at modifying emotional experience and expression. A common definition of emotion regulation encompasses any control processes that are executed to manipulate when, where, how, and which emotions we experience and express (Gross, 1998). Although the term *regulation* may connote an effortful, conscious process, emotion regulation is considered to occur at both automatic and conscious levels of processing. Emotion regulation has both interindividual and intraindividual goals. Most people prefer to feel good and avoid feeling bad and therefore exert control over their emotions to boost positive feelings and minimize negative feelings (e.g., Taylor, 1991). Additionally, when emotions clash with societal norms, people regulate the expression of their emotion in order to avoid negative social consequences. Although emotion regulation most often includes suppressing emotion, it is important to note that emotion regulation may involve the suppression, reduction, or inflation of emotion.

Perhaps the most challenging aspect of studying emotion regulation is measuring whether regulation occurred. If people automatically regulate their emotion, then any emotion response (or lack thereof) may reflect an emotion regulation process. In other words, even if emotion is

expressed, it could reflect an automatic regulation of its magnitude. Most researchers avoid this issue by instructing participants to regulate their emotions through one of two strategies: (1) distracting one's self away from emotional stimuli and (2) reappraising the meaning of emotional stimuli. These studies have provided important information but cannot completely rule out the possibility that the neural regions that support emotion regulation in these studies reflect participants' efforts to follow the experimental instructions and not emotion regulation per se. Therefore it is important to examine naturally occurring emotion regulation processes. But how can researchers do this? Although the topic is not often included in discussions of the neural basis of emotion regulation, research on the spontaneous regulation of the influence of emotion on decision making provides a valuable complement to the studies that explicitly instruct participants to regulate their emotional experience. An examination across these literatures provides an answer to whether different types of emotion regulation strategies (e.g., distraction versus reappraisal, instructed versus spontaneous, regulation of emotional experience versus its influence on subsequent cognition) engage different neural systems. Additionally, these studies provide insight into whether the executive function systems associated with the control of nonemotional stimuli also extend to the control of emotional stimuli.

"Sticking your head in the sand": Distracting attention away from emotional stimuli

One way in which people can control their emotional reactions is to inhibit their attention to events that elicit emotion in the first place. For example, if you are trying to enjoy a nice meal at a restaurant, you might find it annoying when someone at the next table is carrying on a loud conversation on a cell phone. To get rid of your annoyance, you might decide to block out the conversation and ignore it all together. Neuroscientists have examined the neural systems that support the ability to inhibit attention to events that elicit emotion in two ways: distracting attention away from an *anticipated* emotional event and inhibiting attention when distracting emotional information *interferes* with the task at hand.

Anticipation Studies examining the process of distraction when someone is anticipating an emotional event have found the most consistent evidence for the role of the lateral prefrontal cortex (see figure 66.1). These studies typically forewarn participants that emotional events will occur and require the participants to distract their attention away from these events. Participants are explicitly instructed to suppress any thoughts about the emotional events. For example, participants were forewarned that they would be presented with either a negative picture or a neutral picture and were instructed to distract their attention away from these upcoming events by focusing on the experimental context (e.g., "I am lying in a scanner"). Participants who engaged in this distraction process when anticipating a negative emotional picture compared to a neutral picture recruited greater activity in the lateral prefrontal cortex activity (BA 9) and medial prefrontal cortex (BA 8) and exhibited amygdala deactivation (Herwig et al., 2007). In addition to negative pictures, researchers have examined the neural systems that support distraction in the face of upcoming pain. For example, distraction was also associated with lateral prefrontal cortex activation (BA 9/46) when participants anticipated an electric shock, but this activation was generally associated with anticipating an upcoming event regardless of whether it was an electric shock or the absence of shock (Kalisch, Wiech, Hermann, & Dolan, 2006). Together, these studies suggest that frontal regions support the ability to distract attention away from an upcoming emotional event, but it is important to note that the lateral prefrontal cortex has generally been implicated in suppression of nonemotional thoughts (e.g., Mitchell et al., 2007).

Interference Other studies require participants to perform a cognitive task while they simultaneously regulate interference from emotional distractors. To successfully

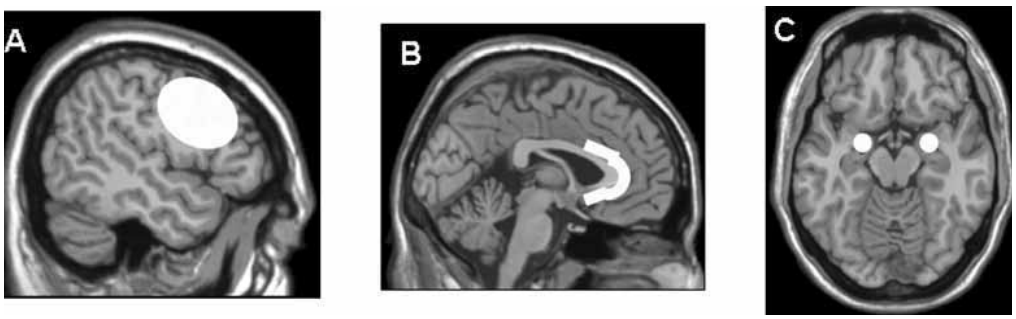


Figure 66.1 Regions associated with distraction: lateral prefrontal cortex (*A*), anterior cingulate cortex (*B*), and amygdala (*C*).

perform the tasks, participants must inhibit their attention to the emotionally distracting information. These studies have found that inhibiting attention to emotional distractors is most consistently associated with (1) increased activation in the anterior cingulate and (2) decreased activation in the amygdala (see figure 66.1). A number of studies have examined participants' ability to distract themselves while receiving pain in the form of a cold pressor task or the application of heat. These studies have found that distracting oneself from physical pain most consistently recruits anterior cingulate cortex (ACC) activity and some other frontal regions (Bantick et al., 2002; Frankenstein, Richter, McIntyre, & Remy, 2001; Wiech et al., 2006). For example, participants' self-reports of negative affect correlated negatively with activation in the ACC (BA 32, BA 24), dorsolateral prefrontal cortex (BA 46), ventrolateral prefrontal cortex (BA 45), and medial prefrontal cortex (BA 10, BA 32) during heat stimulation (Wiech et al., 2006). However, it is unclear whether the activation in the ACC occurs in a consistent subregion. In one study, participants who performed a verbal attention task to distract themselves from a cold pressor on their foot reported less pain and significantly activated the caudal portion of the ACC (BA 32) (Frankenstein et al., 2001). In contrast, another study found that participants who distracted themselves from painful heat reported less pain and significantly activated a rostral portion of the ACC and the orbitofrontal cortex (Bantick et al., 2002).

In addition to studies of regulating interference from pain, researchers have investigated participants' ability to regulate their response to threat cues while performing another task. For example, participants were presented with pairs of house pictures and had to decide whether the same house was depicted across the pictures. The house pictures were flanked by pictures of emotional facial expressions that were either neutral or fearful. The study compared conditions in which the fearful faces were presented either infrequently or frequently. The infrequent condition reflected more interference because participants did not have time to habituate to the fearful cues through satiation. Rostral ACC (BA 32) activated in relation to infrequent fearful distractors. Furthermore, participants' state ratings of anxiety predicted reduced ACC activity and less recruitment of dorsolateral prefrontal cortex (DLPFC) and ventral lateral prefrontal cortex (VLPFC). In other words, regulating away interference from threat cues engaged the ACC, and people who were less successful at regulating their anxiety were the least likely to recruit the ACC, DLPFC, and VLPFC (Bishop, Duncan, Brett, & Lawrence, 2004).

Additionally, several studies have examined the regulation of emotional interference through variations of the Stroop paradigm. For example, participants might be presented with pictures of emotional facial expressions that have emotion words written across the middle. Their task is to identify the emotional content of the facial expression; emotion words that are different from the content of the facial expression represent a condition in which participants have to inhibit their attention to the interfering emotional information. Participants do not tend to make as many errors on this task as they do on the traditional color Stroop task. Instead, emotional interference is indicated by slower reaction times for the interference trials. Research has shown that the presence of emotionally interfering words (as compared to words that match the context of the facial expression) activate dorsomedial prefrontal cortex (BA 8/32), DLPFC, and amygdala (Compton et al., 2003; Etkin, Egner, Peraza, Kandel, & Hirsch, 2006; Egner, Etkin, Gale, & Hirsch, 2008). Participants who successfully regulate the emotional interference significantly activate rostral anterior cingulate (BA 24/32), and activity in this area predicts decreased activation in the amygdala (Etkin et al., 2006; Egner et al., 2008). Modulation of the rostral anterior cingulate was also found in a study that employed an Emotional Counting Stroop paradigm (Whalen et al., 1998). In this study, participants had to count the number of negative or neutral words on the screen. Increased anterior cingulate activation was initially associated with presentation of negative words compared to neutral words. Although no differences in reaction times were found between the negative and neutral conditions in this study, there may have still been emotional interference. Participants may have habituated to the emotional interference, and therefore the effect was diminished after the initial trials (Compton et al., 2003).

Just as there was inconsistency in the subregion of anterior cingulate activation associated with regulating interference from pain, mixed evidence has been found for the selectivity of the subregion of the ACC associated with regulating interference from information that is specifically emotional (rather than any distracting information). On the one hand, one study compared the interference of emotional information with the interference of nonemotional information (e.g., when the gender of a face conflicts with a gender label written across the face) and found that rostral anterior cingulate activation was specific to the interference of emotional information rather than all kinds of information (Egner et al., 2008). However, a functional magnetic resonance imaging (fMRI) study found that the Emotional Stroop task was associated with activation in the caudal portion of the ACC typically found in relation to a traditional Stroop interference effect (Haas, Omura, Constable, & Canli, 2006), and a single-cell recording study found that several cells in this caudal region of the anterior cingulate were specific to the counting Emotional Stroop task (Davis et al., 2005). Although there is no clear explanation for the differences in findings across the studies, one possible difference between the two

studies that found caudal activation and the rest of the research reviewed here is that these studies used positively or negative valenced words that were not descriptions of emotional states; many of the other studies did use emotional state terms. Additionally, one of the studies did not involve the presentation of pictures of emotional facial expressions (Davis et al., 2005).

Reappraisal: "Making lemonade out of lemons"

Another way in which people regulate their emotion is by reappraising events that give rise to emotion. The meaning of the event is changed so that it becomes nonemotional or elicits a different emotion. An important distinction between distraction and reappraisal paradigms is that reappraisal paradigms require participants to decrease or increase their emotions while still attending to the emotional stimuli rather than distracting themselves by looking away or focusing on irrelevant thoughts. In a typical reappraisal experiment, participants receive a fair amount of training on how to reappraise emotional stimuli before they even enter the scanner (e.g., Ochsner, Bunge, Gross, & Gabrieli, 2002; Phan et al., 2005; Urry et al., 2006). Participants are presented with emotional stimuli and instructed to generate an alternative interpretation of the stimuli so that the salient emotional meaning is changed. For example, a negative scenario such as a baby covered in electrodes while receiving medical treatment can be reappraised as a positive event because the treatment is necessary for a healthy recovery. Researchers can then compare neural activity associated with reappraising the emotional meaning of stimuli to neural activity associated with reacting to the most salient emotional meaning of stimuli.

In contrast to the research on distraction, studies of reappraisal use less varied paradigms. These studies mostly focus on reappraisal as a means of suppressing a negative emotional experience. The bulk of these studies focus on reducing negative emotion through the reappraisal of negative pictures from the International Affective Picture Set (IAPS) (Lang, Bradley, & Cuthbert, 1995) in young healthy adults (e.g., Banks, Eddy, Angstadt, Nathan, & Phan, 2007; Eippert et al., 2007; Kim & Hamann, 2007; Ochsner et al., 2002; Ohira et al., 2006; Phan et al., 2005; Schaefer et al., 2002) or older adults (Urry et al., 2006), and some have examined reappraisal of negative IAPS that represent moral violations (e.g., Harenski & Hamann, 2006). A few studies have examined the neural basis of reducing negative emotion through the reappraisal of sad films in adults (e.g., Lévesque et al., 2003) and children (e.g., Lévesque et al., 2004). However, there are some exceptions to the focus on the reduction of negative emotion. Some studies have examined the reduction of positive emotion through reappraisal of positive stimuli (e.g., Beauregard, Lévesque, & Bourgouin, 2001; Kim & Hamann, 2007) or the inflation of emotion (e.g., Kim & Hamann, 2007; Ochsner et al., 2004).

ERP research suggests that reappraisal happens as early as 200 ms after the onset of the emotional stimulus and is sustained for at least 2 s (Hajcak & Nieuwenhuis, 2006). Whereas the distraction research most consistently found activation in the ACC and deactivation in the amygdala, fMRI and PET research suggests that the reappraisal process is governed by (1) *increased* activation in frontal regions such as lateral prefrontal cortex, medial and lateral orbitofrontal cortex, and anterior cingulate/dorsomedial prefrontal cortex and (2) *decreased* activation in the amygdala (see figure 66.2).

Lateral Prefrontal Cortex A number of studies have found that reappraising stimuli to reduce emotional experience is associated with activity in the ventrolateral and lateral prefrontal cortex. For example, studies of healthy, young adult participants as they reappraise negative pictures or films have found significant increases in activation in the ventrolateral prefrontal cortex (BA 46/10: Ochsner et al. 2002; BA 44/46: Kim & Hamman, 2007; Phan et al., 2005) and the lateral prefrontal cortex (Banks et al., 2007; Eippert et al., 2007; Kim & Hamann, 2007; Lévesque et al., 2003;

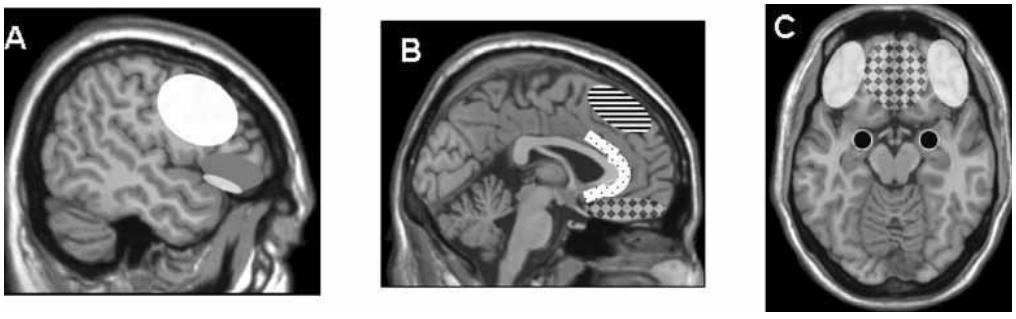


Figure 66.2 Regions associated with reappraisal. Dorsolateral (white in *A*) and ventrolateral (dark gray in *A*) prefrontal cortex. Medial (checkerboard in *B*, *C*) and lateral (light gray in *A*, *C*) orbitofrontal cortex, dorsomedial prefrontal cortex (striped in *B*), anterior cingulate (dotted in *B*), and amygdala (black with white outline in *C*).

Ohira et al., 2006). Similar regions of activation have been associated with the reappraisal of negative stimuli in older adult populations (BA 9, BA 44: Urry et al., 2006) and in children (BA 47, BA 9: Lévesque et al., 2004). Additionally, a study of individual differences in rumination suggests that participants who are likely to have poor ability to regulate their emotion (i.e., report high levels of dispositional rumination) do not strongly recruit the lateral prefrontal cortex when instructed to reappraise negative stimuli (Ray et al., 2005). Together, these studies suggest a consistent role of the lateral prefrontal cortex, a region that is associated with control of nonemotional responses, in the process of reappraisal. Furthermore, activation in lateral prefrontal cortex is reduced for individuals who struggle to reappraise emotional stimuli (e.g., ruminators).

Although many of the reappraisal studies explicitly instruct participants to reappraise stimuli, two studies suggest that regions of lateral prefrontal cortex may also be associated with the automatic regulation of negative emotion (Cheng et al., 2007; Haas et al., 2006). In one study, participants were presented with pictures of facial expressions that were fearful, happy, sad, or neutral. Participants were instructed to rate the gender of the person in each picture. On the basis of previous research suggesting that individuals who are high in agreeableness are automatically motivated to regulate away threat signals (Tobin, Graziano, Vanman, & Tassinary, 2000), the authors of this fMRI study reasoned that highly agreeable individuals would be likely to automatically regulate their response to fearful faces but not happy or sad faces. The results were consistent with this hypothesis: self-reported ratings of agreeableness were positively correlated with lateral prefrontal activation (BA 9) in relation to the fearful face trials but not related to the happy or sad face trials (Haas et al., 2006). In a second study, physicians who practice acupuncture and acupuncture novices were presented with short film clips of patients receiving acupuncture (Cheng et al., 2007). In comparison to novices, who activated neural regions associated with empathic pain, the physicians activated lateral prefrontal cortex. The authors of the study suggest that the lateral prefrontal cortex activity reflects the physicians' spontaneous tendency to regulate away any empathic pain by focusing on how the treatment will help the patient (Cheng et al., 2007).

Although the bulk of studies examine the reappraisal of negative emotion, a few studies have found lateral prefrontal activity in relation to reappraisal in other kinds of paradigms and, in some cases, more complex emotions. For example, bilateral ventrolateral prefrontal cortex (BA 47) activity is associated with reappraising moral emotion pictorial stimuli (Harenski & Hamann, 2006), and lateral prefrontal cortex (BA 44) is associated with reappraising erotic emotional stimuli (Beauregard et al., 2001). The relationship between lateral prefrontal cortex activity and reappraisal also extends to reappraisals aimed at intensifying emotional experience. Lateral prefrontal cortex activity (BA 9 and BA 44) has been found in association with increasing negative emotional experience in young adults (Ochsner et al., 2004) and elderly adults (Urry et al., 2006) as well as increasing positive emotional experience in young adults (Kim & Hamann, 2007).

Medial and Lateral Orbitofrontal Cortex A number of reappraisal studies have also found significant medial and lateral orbitofrontal cortex activation. For example, the reappraisal of negative pictures or sad films is associated with medial orbitofrontal activity (BA 11) in children (Lévesque et al., 2004) and adults (Banks et al., 2007; Eippert et al., 2007; Kim & Hamann, 2007; Lévesque et al., 2003; Ochsner et al., 2004; Phan et al., 2005). The reappraisal of positive IAPS is also associated with bilateral orbitofrontal cortex activity (Kim & Hamann, 2007). Furthermore, not only is medial orbitofrontal cortex activity associated with the reappraisal condition in neuroimaging studies, but there is also some evidence that this activity correlates with peripheral nervous system indicators of emotional suppression. One neuroimaging study found increased medial orbitofrontal activity (BA 11) in relation to the reappraisal of negative emotional stimuli and that this activity correlates with increased skin conductance response (SCR) (Ohira et al., 2006). Increased skin conductance has been associated with emotional suppression (e.g., Gross & Levenson, 1993).

Anterior Cingulate and Medial Prefrontal Cortex A number of studies have also found that reappraisal is associated with anterior cingulate activity. For example, anterior cingulate activity (BA 24/32) is associated with the reappraisal of negative IAPS (Banks et al., 2007; Eippert et al., 2007; Kim & Hamann, 2007; Ochsner et al., 2002; Phan et al., 2005) and moral emotional pictures (Harenski & Hamann, 2006) in adults and the reappraisal of sad films in children (Lévesque et al., 2004). Anterior cingulate (BA 32) activity is also involved in reducing positive emotion through the reappraisal of positive stimuli (Beauregard et al., 2001). In addition to reducing emotion, the ACC has been implicated in increasing emotion through reappraisal of both negative and positive pictures (Kim & Hamann, 2007).

A number of reappraisal studies have also found significant activation in regions of the dorsomedial prefrontal cortex that are broadly adjacent to the anterior cingulate activation. For example, regions of BA 8/9 show significant activation in relation to the reappraisal of negative IAPS (Banks et al., 2007; Kim & Hamann, 2007; Ochsner et al., 2002; Phan et al., 2005). Additionally, BA 9/10 activity has been associated with the reappraisal of sad films (Lévesque et al., 2003) and moral emotional stimuli

(Harenski & Hamann, 2006). Finally, the left medial prefrontal cortex (BA 9) has been associated with increasing negative emotion through reappraisal of negative IAPS (Ochsner et al., 2004).

Amygdala Reappraising emotional stimuli modulates amygdala activity. For example, a decrease in amygdala activity is associated with reappraising negative IAPS in young (Banks et al., 2007; Eippert et al., 2007; Ochsner et al., 2004; Phan et al., 2005) and elderly adults (Urry et al., 2006) and the reappraisal of positive erotic stimuli (Beauregard et al., 2001). Similarly, amygdala activity increases in relation to reappraisal aimed at (1) inflating *negative* emotional experience in young (Eippert et al., 2007; Ochsner et al., 2004; Schaefer et al., 2002) and elderly adults (Urry et al., 2006) and (2) inflating *positive* emotional experience in young adults (Kim & Hamann, 2007).

Relation of Frontal Regions and Amygdala Some studies suggest that the relative activation of frontal regions and relative decrease amygdala activity may be related. One fMRI study required female participants to reappraise negative pictures and found that reappraisal was associated with left ventrolateral prefrontal cortex (BA 46/10) activation correlated with decreased right amygdala activation (Ochsner et al., 2002). Another fMRI study that included both male and female participants and involved reappraisal of negative pictures found that reappraisal was associated with a negative correlation between right ventrolateral prefrontal cortex activation (BA 47) and left amygdala activation (Phan et al., 2005). Although these studies found a general relationship between the frontal lobes and amygdala, other studies have found individual differences in the frontal-amygdala coupling when participants are reappraising negative IAPS (Urry et al., 2006) and the strength of the frontal-amygdala coupling predicts individual differences in self-reported regulation success (i.e., a reduction in negative emotion) (Banks et al., 2007). In contrast to the studies reviewed above, numerous studies examining reappraisal of negative IAPS or films do not report a significant correlation between frontal activity and amygdala deactivation. It is possible that this relationship was not tested and, therefore, the failure to report the negative correlation might not indicate a nonreplication.

Regulating the influence of emotional expression on decision making

Distraction and reappraisal studies often explicitly instruct participants to regulate their emotional experiences in relation to anticipated or present emotional stimuli. A remaining question is whether the neural regions associated with emotional regulation in these studies extend to spontaneous emotion regulation. Additionally, these studies focus on the regulation of emotional experience and do not address the regulation of emotion's expression on complex cognition such as decision making. To address these questions, it is necessary to explore research literatures that may not always be included in reviews of the neural basis of emotion regulation. In particular, an important avenue for addressing these questions is research examining situations in which participants spontaneously regulate the influence of emotion on their decision making.

Much of the research that examines the spontaneous regulation of emotional influences on decision making are designed from the perspective of economic models of decision making. Economic models of decision making assume that people should make self-interested decisions, that is, decisions that maximize their financial gains and minimize their financial losses (e.g., Kahneman & Tversky, 2000). However, people may sometimes react emotionally to some aspect of a financial decision that is irrelevant, and if they follow these emotions, they make decisions that sacrifice monetary gain. Therefore, it is possible to examine the spontaneous regulation of emotional expression on decision making by creating situations in which participants' self-interest competes with their emotional reactions. Studies adopting this approach have most often found that spontaneous regulation of emotional expression on decision making is associated with dorsolateral prefrontal cortex, medial and lateral orbitofrontal cortex, and anterior cingulate activity (see figure 66.3).

The Ultimatum Game is one paradigm that measures differences in gambling decisions that reflect the influence of an initial but irrelevant emotional reaction or the regulation of that emotional reaction. As a responder in the Ultimatum Game, participants must decide whether to accept or reject another player's suggestion about how to split a sum of money provided by the experimenters. Some offers are fair (e.g., very close to 50% for each person), but others are unfair (e.g., 80% for the player and 20% for the participant). If the responder accepts the offer, the participants receive the money according to the terms of the offer. If the responder rejects the offer, then neither player receives any money. Participants tend to have a negative emotional reaction when presented with an unfair offer. However, from an economic perspective, participants should not let their negative emotional reaction lead them to reject the unfair offer. Even if it is unfair, they can gain some money instead of sacrificing any chance of monetary gain. This is particularly true if the responder plays several rounds of the game with different participants. In this case, there is no rationale for rejecting offers such as worries about communicating a willingness to take less than an equal share. The responder will never interact with that same player again.

In the first neural study of this paradigm, the consideration of unfair offers was associated with ACC, dorsolateral

FIGURE 66.3 Regions associated with regulating emotional expression on decision making. Lateral prefrontal cortex (white in *A*), medial (checkerboard in *B*, *C*) and lateral (light gray in *A*, *C*) orbitofrontal cortex, and anterior cingulate cortex (dotted *B*).

prefrontal cortex, and insula activity (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Insula activity has often been associated with negative emotions such as disgust, anger, pain, and distress suggesting that this activity may have reflected these emotional reactions to the unfair offers. Furthermore, increased insula activity during consideration of an unfair offer predicted the likelihood that the offer would be rejected. Although dorsolateral prefrontal cortex activity did not directly correlate with the acceptance of unfair offers, the authors of the study noted that participants tended to accept unfair offers if they activated their dorsolateral prefrontal cortex more strongly than their insula when considering unfair offers. They suggest that the differential activation between insula and dorsolateral prefrontal cortex reflects a difference in decision making driven by an emotion reaction supported by the insula in comparison to decision making driven by a cognitive control system supported by the dorsolateral prefrontal cortex.

Although it would be tempting to interpret the dorsolateral prefrontal activation to indicate a reappraisal of the unfair offers as suggested by the studies of reappraisal, two studies using transcranial magnetic stimulation (TMS) suggest that it is somewhat more complicated. Unfair offers violate norms of fairness, and participants' initial reaction is to reject the offer. Therefore it might be expected that the dorsolateral prefrontal cortex supports the reappraisal of the unfair offer so that participants are able to regulate their initial negative reaction and focus on the choice between earning some money compared to no money. If the dorsolateral activation is needed to generate a reappraisal of the unfair offer so that it becomes appealing, then disruption of this region through TMS should increase rejection rates of unfair offers. However, two TMS studies have found that disruption of this region is associated with higher rates of acceptance of unfair offers compared to sham stimulation (Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Knoch et al., 2007). In light of the replication of an unexpected finding, the authors of the TMS studies suggest that these results could indicate that the dorsolateral prefrontal cortex is actually needed to implement fairness norms when they conflict with the temptation of selfish interests. In other words, the dorsolateral prefrontal cortex is not needed to suppress negative reactions to unfair offers but rather to suppress the desire for monetary gain when giving into that desire could be costly for your social standing. Social norms of fairness dictate that people should not have to accept unfair offers. Therefore deciding to accept unfair offers is not the rational choice, as it might indicate that you are willing to receive less than your share in future situations.

The role of the dorsolateral prefrontal cortex in implementing social norms has also been suggested in neural research on moral judgments. For example, participants had to make decisions about what course of action to take for various moral dilemmas. The most difficult dilemmas created a conflict between the course of action suggested by an initial emotional response and the course of action suggested by utilitarian values (e.g., what is good for the group rather than the individual) (Greene, Nystrom, Engell, Darley, & Cohen, 2004). For example, one dilemma placed participants in a situation in which they were part of a group of people hiding from soldiers in wartime. The participants were told that their baby was crying and that this could cause discovery of the whole group, which would mean certain death. Participants then had to decide whether they would smother the baby to stop the crying in order to save the group. Difficult moral judgments were associated with ACC and dorsolateral prefrontal cortex activity. The authors of the study suggest that the ACC helps to detect the conflict presented by these dilemmas and the dorsolateral prefrontal cortex supports the selection of the utilitarian response.

Regulation of the clash between self-interest and fairness norms may also be supported by orbitofrontal cortex. For example, one fMRI study required participants to determine how much of a sum of money they were willing to share with another participant (Spitzer, Fischbacher, Herrnberger, Groen, & Fehr, 2007). Participants made these decisions in two conditions. In a control condition, the offer was seen only by the participant and the other player. In the punishment condition, a third person had the power to take away

the money if the offer was perceived as unfair. The study found that dorsolateral prefrontal cortex and bilateral lateral orbitofrontal cortex activation were associated with offering more money in the punishment condition than in the control condition. Furthermore, a high score on a measure of Machiavellianism predicted bilateral orbitofrontal cortex activation in the punishment condition. Machiavellianism measures the extent to which individuals are willing to compromise their behavior as long as it helps them to achieve their ultimate goal. These results suggest that individuals who are most likely to adhere to norms of fairness only as a means of achieving their self-interest in gaining money are those individuals who tend to activate frontal regions associated with emotion regulation.

Other research has implicated the orbitofrontal cortex in regulating the influence of emotion on decision making. For example, the presentation of equivalent gambling decisions that are framed as guaranteed gains or losses measures people's tendency to let emotion guide their decisions in favor of controlling its influence. A host of research has shown that people want to avoid loss and therefore will choose to gamble when faced with a guaranteed loss even if the money at stake is the same as a bet that is framed as a guaranteed gain. In other words, if people are endowed with 50 dollars and given a choice between a guaranteed loss of 30 dollars or gambling it all, people will choose to gamble. In contrast, if given a choice between a guaranteed gain of 20 dollars and gambling it all, people will choose to keep the 20 dollars. However, the guaranteed money in both bets is equivalent. A guaranteed loss elicits a negative emotional response and makes people focus on any option that will help them to avoid the guaranteed loss. However, acting on emotion is detrimental because, as was mentioned above, participants are not making decisions based on their actual monetary consequences. An fMRI study of equivalent bets framed as gains or losses found that participants who permitted their betting behavior to be swayed by the loss frames showed significant amygdala activation and those who bet according to monetary amount (rather than frame) showed increased medial and lateral orbitofrontal cortex activation (DeMartino, Kumaran, Seymour, & Dolan, 2006).

Although some previous research suggests that the orbitofrontal cortex is important for regulating the influence of emotion on decision making, it is possible that the orbitofrontal cortex activation reflected the regulation of the emotional experience itself because the emotional reactions occurred simultaneously with the decision. However, another fMRI study disentangled the regulation of emotional experience from the regulation of emotional influence on decision making and found that lateral orbitofrontal cortex and insula activity are associated with the regulation of emotional influence on decision making (Beer, Knight, & D'Esposito, 2006). Participants were presented with negative and neutral cues and were asked to either rate the valence of the cues or ignore them. After each cue, participants were presented with the odds and payoff information for a roulette bet and were asked to select a monetary amount to wager. Previous research has shown that negative moods make participants less willing to risk money. Therefore if participants fail to successfully ignore the negative cues, then they will risk less when primed with a negative cue compared to a neutral cue. Participants who successfully regulated the influence of these negative primes on their betting behavior were most likely to engage the lateral orbitofrontal cortex and insula (Beer, Knight, & D'Esposito, 2006).

Assessment of neuroimaging research of emotion regulation

The chapter began by posing a number of questions regarding the neural basis of emotion regulation and sought to address them by comparing and contrasting research on the emotion regulation processes of distraction, reappraisal, and controlling the influence of emotional expression on decision making. These bodies of research suggest that although there is overlap between the neural regions associated with emotion regulation, there are systems that appear more consistently in relation to particular emotion regulation strategies. Furthermore, the neural systems that are associated with instructed and spontaneous emotion regulation share some similarities. Finally, there is not strong evidence for a "special" executive function system that controls emotion. The neural regions associated with emotion regulation are regions that are also associated with the control of nonemotional information.

Commonalities Across Emotion Regulation Strategies

All three of the emotion regulation processes that we consider in this chapter were associated with the lateral prefrontal cortex and ACC. These regions have long been associated with executive function; therefore it is not surprising that they would also support control over emotional processes. Although very little research has experimentally manipulated emotion regulation in human lesion populations, there is some evidence that lesions to the lateral prefrontal cortex and ACC disrupt emotion regulation abilities. For example, patients with lateral prefrontal cortex damage have reported increases in depression and anxiety, and patients with ACC damage have been associated with both increased and decreased emotionality, suggesting that these patients' ability to engage in emotion regulation is impaired (see Beer, 2007a, for a review). The role of lateral prefrontal cortex in emotion regulation appears to extend across both instructed regulation efforts and spontaneous regulation of emotion. In addition

to the research on the spontaneous regulation of emotional influences on cognition, it is the case that individual differences associated with spontaneous tendencies toward emotion regulation are associated with lateral prefrontal cortex activity. For example, lateral prefrontal cortex activity is associated with the regulation efforts of participants whose high levels of agreeableness motivate them to regulate their attention to threat cues (Haas et al., 2006) and with physicians who spontaneously regulate empathic pain when watching patients receive acupuncture (Cheng et al., 2007). Additionally, the lateral prefrontal cortex is recruited less strongly by individuals who are poor at emotion regulation, such as highly anxious subjects (Bishop et al., 2004) and highly ruminative subjects (Ray et al., 2005).

Differences across Emotion Regulation Strategies However, there were differences across the emotion regulation strategies as well. For example, the studies suggest that distraction is supported by the inverse relation between the ACC and the amygdala (e.g., Etkin et al., 2006; Egner et al., 2008), whereas reappraisal is supported by the inverse relation between the lateral orbitofrontal cortex and the amygdala (e.g. Banks et al., 2007; Ochsner et al., 2002; Phan et al., 2005). These relationships parallel findings in nonhuman animal research. For example, the process of extinction requires animals to no longer respond to the emotional association of a stimulus. This process is arguably similar to the process of distraction in which people suppress their emotional response. A similar inverse relationship between a region in the rat medial prefrontal cortex and the amygdala is associated with extinction (see Quirk & Beer, 2006, for a review). Similarly, an inverse coupling between the lateral orbitofrontal cortex and the amygdala is associated with reversal learning in rats (see Holland & Gallagher, 2004, for a review). Studies of reversal learning in nonhuman animals suggest that the amygdala stores emotional associations and the lateral orbitofrontal cortex supports the flexible use of these associations (e.g., Jones & Mishkin, 1972).

Additionally, the neural systems that are associated with distraction and reappraisal are more widespread than are those found in relation to the spontaneous regulation of emotion's influence on decision making. Although there are a number of differences between these bodies of research, one possibility is that spontaneous regulation of emotion relies on fewer neural structures when compared to effortful regulation. It may also be that similar systems support spontaneous emotion regulation regardless of the specific strategy; the additional neural activity found in distraction and reappraisal studies might reflect the participants' efforts to follow a set of instructions. The neural distinction between emotion regulation that is instructed versus spontaneous emotional regulation is illustrated in the dissociation between the ability of patients with orbitofrontal lesions to comply with instructions to regulate their response to emotional film clips (Beer, 2007b) and their failure to regulate their emotional self-disclosures in everyday social situations (Beer, Heerey, Keltner, Scabini, & Knight, 2003, Beer, John, Scabini, & Knight, 2006).

A Unique Executive Function System for Emotion Regulation? Some emotion regulation research has examined the possibility that there could be distinct neural regions that support the control of emotional processes when compared to nonemotional processes. For example, some research has suggested that an Emotional Stroop task recruits a rostral portion of the ACC that is distinct from the caudal portion associated with traditional Stroop tasks (e.g., Egner et al., 2008). In contrast, other research has found caudal anterior cingulate involvement in Emotional Stroop tasks (e.g., Haas et al., 2006; Davis et al., 2005). Although it seems most parsimonious to propose a general control system that can be recruited to handle both emotional and nonemotional information, emotion regulation is a particularly difficult task. By definition, emotional stimuli are highly salient and therefore more difficult to control. Therefore any differences in the neural regions associated with emotion regulation might not indicate a special system for the control of emotion but instead reflect differences in difficulty.

Future directions

In a short time, the neural research on emotion regulation has flourished, but there are also a number of fruitful avenues to broaden this body of research. The most important direction will be studies that are aimed at understanding how these neural systems interact to support emotion regulation and their relationship to peripheral nervous system indicators of emotion regulation. The large body of extant work provides a strong foundation for hypothesized activations; therefore it will be most beneficial to move past the identification of regions and understand how they interact. These studies will be beneficial in making progress on understanding the psychological function of each region in the emotion regulation process. At the moment, most models relating the anatomy to psychological process can only speculate on the psychological contribution of neural regions on the basis of studies of cognitive control (e.g., Ochsner & Gross, 2005). For example, the debate about the role of the ACC in detecting conflict or resolving conflict extends to research on emotion regulation, for example, detecting conflict (Bishop et al., 2004) and resolving conflict (Etkin et al., 2006). More careful research will permit stronger conclusions about the specific function of the ACC in relation to different emotion regulation strategies. Additionally, some studies have begun

to examine how neural activations relate to indicators of emotional regulation in the peripheral nervous system, and it will be helpful to continue understanding how these systems interact in order to gain a fuller picture of the process of emotion regulation.

More lesion work is needed in the study of emotion regulation. Almost no experimental manipulations of emotion regulation have been conducted in human lesion populations (see Beer, 2007a, for a review). An obvious place to begin is to understand the emotion regulation capabilities of patients with dorsolateral prefrontal cortex damage and orbitofrontal cortex damage because these regions have been identified in the neural research and patients with selective damage to these regions are more likely to exist than patients with damage to other relevant regions. Studies of distraction, reappraisal, and regulating emotional influence on decision making are needed to complement the work that has been done with neuroimaging techniques such as PET, fMRI, and TMS.

Future studies should also aim to directly compare the strategies of distraction and reappraisal. Although a number of overlapping regions are associated with both strategies, the two strategies may be distinguished by the frontal modulation of amygdala activity (ACC for distraction and orbitofrontal cortex for reappraisal). However, the overlap might occur because participants tend to distract themselves somewhat while reappraising and vice versa. A stronger test of the regions that might be more strongly activated in relation to a specific strategy will be to directly compare these strategies within the same study.

Future studies should include more varied emotional stimuli and samples that are balanced for gender. Much of the current work on reappraisal involves reducing negative emotional reactions to IAPS pictures. This paradigm has been helpful in permitting comparison of findings across laboratories, but it will also be helpful to extend this research to different emotional stimuli. Some studies have found that participants exhibit individual differences in the strength of their emotional reactions to this standardized set. Therefore, researchers sometimes have to specifically examine the trials that received the hypothesized emotional ratings by participants in order to detect neural activity (e.g., Ochsner et al., 2002). The use of films could elicit stronger and more uniform emotional responses and make it easier to elicit a wider range of emotions. Investigating the regulation of emotions that are more complex will be particularly important for understanding how the frontal lobes support regulation through the modulation of other neural structures. The extant work suggests that an inverse relationship between amygdala activity and regions within the frontal lobe is important for the regulation of emotion (e.g., ACC in the case of distraction and lateral orbitofrontal cortex in the case of reappraisal). However, the focus on the modulation of the amygdala might have arisen because many of these studies focus on the regulation of reactions to negative pictures that are likely to elicit basic emotions associated with amygdala function such as fear, anger, sadness, and disgust. The investigation of more complex emotions, often called *social emotions* or *moral emotions* (e.g., pride, embarrassment, shame, guilt), could reveal different inverse relationships in cases of regulation because the generation of these emotions tend to be associated with activity in the frontal lobes (see Beer, 2007c, for a review).

Additionally, samples that include both men and women are needed for future studies of distraction and reappraisal. As was mentioned above, the neural regions associated with regulating emotional experience, whether through distraction or reappraisal, tend to be more widespread than the neural regions associated with regulating the influence of emotion on decision making. However, many of the distraction and reappraisal studies include only female participants, whereas the studies of regulated decision making typically include a more even gender balance.

Conclusion

Although emotion is a useful deviation from our homeostatic baseline, we need to flexibly regulate its experience and expression. Just as the frontal lobes permit the control of nonemotional information, they are recruited for the task of regulating our emotions. Although a strong foundation of research has identified neural regions that typically support emotion regulation, more research is needed to fully understand how these neural regions work together to allow us to feel good most of the time and avoid social banishment when those inconvenient emotions clash with societal norms.

REFERENCES

Banks, S. J., Eddy, K. T., Angstadt, M., Nathan, P. J., Phan, K. L. (2007). Amygdala-frontal connectivity during emotion regulation. *Soc. Cogn. Affect. Neurosci.*, *2*, 303–312.

Bantick, S. J., Wise, R. G., Ploghaus, A., Clare, S., Smith, S. M., & Tracey, I. (2002). Imaging how attention modulates pain in humans using functional MRI. *Brain*, *125*, 310–319.

Beauregard, M., Lévesque, J., & Bourgouin, P. (2001). Neural correlates of conscious self-regulation of emotion. *J. Neurosci.*, *21*, 1–6.

Beer, J. S. (2007a). Insights into emotion regulation from neuropsychology. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 69–86). New York: Guilford.

Beer, J. S. (2007b). The importance of emotion-cognition interactions for social adjustment: Insights from the orbitofrontal cortex. In E. Harmon-Jones & P. Winkielman (Eds.), *Social neuroscience: Integrating biological and psychological explanations of social behavior* (pp. 15–30). New York: Guilford.

Beer, J. S. (2007c). Neural systems for self-conscious emotions and their underlying appraisals. In J. L. Tracy, R. W. Robins, &

J. P. Tangney (Eds.), *The handbook of self-conscious emotions* (pp. 53–67). New York: Guilford.

Beer, J. S., John, O. P., Scabini, D., & Knight, R. T. (2006). Orbitofrontal cortex and social behavior: Integrating self-monitoring and emotion-cognition interactions. *J. Cogn. Neurosci.*, *18*, 871–880.

Beer, J. S., Heerey, E. H., Keltner, D., Scabini, D., & Knight, R. T. (2003). The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *J. Pers. Soc. Psychol.*, *85*, 594–604.

Beer, J. S., Knight, R. T., & D'Esposito, M. (2006). Integrating emotion and cognition: The role of the frontal lobes in distinguishing between helpful and hurtful emotion. *Psychol. Sci.*, *17*, 448–453.

Bishop, S., Duncan, J., Brett, M., & Lawrence, A. D. (2004). Prefrontal cortical function and anxiety: Controlling attention to threat-related stimuli. *Nat. Neurosci.*, *7*, 184–188.

Cheng, Y., Lin, C. P., Liu, H. L., Hsu, Y. Y., Lim, K. E., Hung, D., et al. (2007). Expertise modulates the perception of pain in others. *Curr. Biol*, *17*, 1708–1713.

Compton, R. J., Banich, M. T., Mohanty, A., Milham, M. P., Herrington, J., & Miller, G. A. (2003). Paying attention to emotion: An fMRI investigation of cognitive and emotional Stroop tasks. *Cogn. Affect. Behav. Neurosci.*, *3*, 81–96.

Davis, K. D., Taylor, K. S., Hutchison, W. D., Dostrovsky, J. O., McAndrews, M. P., Richter, E. O., et al. (2005). Human anterior cingulate cortex neurons encode cognitive and emotional demands. *J. Neurosci.*, *25*, 8402–8406.

DeMartino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision making in the human brain. *Science*, *313*, 684–687.

Egner, T., Etkin, A., Gale, S., & Hirsch, J. (2008). Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. *Cereb. Cortex*, *18*, 1475–1484.

Eippert, F., Veit, R., Weiskopf, N., Erb, M., Birbaumer, N., & Anders, S. (2007). Regulation of emotional responses elicited by threat-related stimuli. *Hum. Brain Mapping*, *28*, 409–423.

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., & Hirsch, J. (2006). Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron*, *51*, 871–882.

Frankenstein, U. N., Richter, W., McIntyre, M. C., & Remy, F. (2001). Distraction modulates anterior cingulate gyrus activations during the cold pressor test. *NeuroImage*, *14*, 827–836.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Rev. Gen. Psychol.*, *2*, 271–299.

Gross, J. J., & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behavior. *J. Pers. Soc. Psychol.*, *64*, 970–986.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389–400.

Haas, B. W., Omura, K., Constable, R. T., & Canli, T. (2006). Interference produced by emotional conflict associated with anterior cingulate activation. *Cogn. Affect. Behav. Neurosci.*, *6*, 152–156.

Hajcak, G., & Nieuwenhuis, S. (2006). Reappraisal modulates the electrocortical response to unpleasant pictures. *Cogn. Affect. Behav. Neurosci.*, *6*, 291–297.

Harenski, C. L., Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*, *30*, 313–324.

Herwig, U., Baumgartner, T., Kaffenberger, T., Bruhl, N., Kottlow, M., Schreiter-Gasser, U., et al. (2007). Modulation of anticipatory emotion and perception processing by cognitive control. *NeuroImage*, *37*, 652–662.

Holland, P. C., & Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy. *Curr. Opin. Neurobiol.*, *14*, 148–155.

Jones, B., & Mishkin, M. (1972). Limbic lesions and the problem of stimulus-reinforcement associations. *Exp. Neurol.*, *36*, 362–377.

Kahneman, D., & Tversky, A. (2000). *Choices, values, and frames*. New York: Cambridge University Press.

Kalisch, R., Wiech, K., Herrmann, K., & Dolan, R. J. (2006). Neural correlates of self-distraction from anxiety and a process model of cognitive emotion regulation. *J. Cogn. Neurosci.*, *18*,1266–1276.

Kim, S. H., & Hamann, S. (2007). Neural correlates of positive and negative emotion regulation. *J. Cogn. Neurosci.*, *19*, 776–798.

Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., & Fehr, E. (2007). Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. *Cereb. Cortex*, *18*, 1987–1990.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*, 329–332.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1995). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Gainesville: University of Florida.

Lévesque, J., Eugène, F., Joanette, Y., Paquette, V., Mensour, B., Beaudoin, G., Leroux, J.-M., Bourgouin, P., & Beauregard, M. (2003). Neural circuitry underlying voluntary suppression of sadness. *Biol. Psychiatry*, *53*, 502–510.

Lévesque, J., Joanette, Y., Mensour, B., Beaudoin, G., Leroux, J. M., Bourgouin, P., et al. (2004). Neural basis of emotional self-regulation in childhood. *Neuroscience*, *129*, 361–369.

Mitchell, J. P., Heatherton, T. F., Kelley, W. M., Wyland, C. L., Wegner, D. L., & Macrae, C. M. (2007). Separating sustained from transient aspects of cognitive control during thought suppression. *Psychal. Sci.*, *18*, 292–297.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: An fMRI study of the cognitive regulation of emotion. *J. Cogn. Neurosci.*, *14*, 1215–1229.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends Cogn. Sci.*, *9*, 242–249.

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., et al. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage, 23*, 483–499.

Ohira, H., Nomura, M., Ichikawa, N., Isowa, T., Iidaka, T., Sato, A., (2006). Association of neural and physiological responses during voluntary emotion suppression. *NeuroImage*, *29*, 721–733.

Phan, K. L., Fitzgerald, D. A., Nathan, P. J., Moore, G. J., Uhde, T. W., & Tancer, M. E. (2005). Neural substrates for voluntary suppression of negative affect: A functional magnetic resonance imaging study. *Biol. Psychiatry*, *57*, 210–219.

Quirk, G. J., & Beer, J. S. (2006). Prefrontal involvement in the regulation of emotion: Convergence of rat and human studies. *Curr. Opin. Biol.*, *16*, 723–727.

Ray, R. D., Ochsner, K. N., Cooper, J. C., Robertson, E. R., Gabrieli, J. D. E, & Gross, J. J. (2005). Individual differences in trait rumination and the neural systems supporting cognitive reappraisal. *Cogn. Affect. Behav. Neurosci.*, *5*, 156–168.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision making in the ultimatum game. *Science*, *300*, 1755–1758.

Schaefer, S. M., Jackson, D. C., Davidson, R. J., Aguirre, G. K., Kimberg, D. Y., & Thompson-Schill, S. (2002). Modulation of amygdalar activity by the conscious regulation of negative emotion. *J. Cogn. Neurosci.*, *14*, 913–921.

Spitzer, M., Fischbacher, U., Herrnberger, B., Groen, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, *56*, 185–196.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychol. Bull.*, *110*, 67–85.

Tobin, R. M., Graziano, W. G., Vanman, E. J., & Tassinary, L. G. (2000). Personality, emotional experience, and efforts to control emotions. *J. Pers. Soc. Psychol.*, *79*, 656–669.

Urry, H. L., van Reekum, C. M., Johnstone, T., Kalin, N. H., Thurow, M. E., Schaefer, H. S., et al. (2006). Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *J. Neurosci*, *26*, 4415–4425.

Whalen, P. J., Bush, G., McNally, R. J., Wilhelm, S., McInerney, S. C., Jenike, M. A., et al. (1998). The emotional counting Stroop paradigm: A functional magnetic resonance imaging probe of the anterior cingulated affective division. *Biol. Psychiatry*, *44*, 1219–1228.

Wiech, K., Kalisch, R., Weiskopf, N., Pleger, B., Stephan, K. E., & Dolan, R. J. (2006). Anterolateral prefrontal cortex mediates the analgesic effect of expected and perceived control over pain. *J. Neurosci.*, *26*, 11501–11509.

Wilson, E. G. (2008). Against happiness: In praise of melancholy. New York: Farrar, Straus & Giroux.

67 Sharing the Emotions of Others: The Neural Bases of Empathy

TANIA SINGER AND SUSANNE LEIBERG

ABSTRACT With the emergence of social neuroscience, researchers have started to investigate our ability, known as empathy, to share other people's feelings. After defining empathy and delineating the difference between empathy and related concepts such as sympathy, cognitive perspective taking, and emotional contagion, we present a neuroscientific account of empathy and provide both peripheral and central neurophysiological evidence for it. We then discuss the central role of insular cortex in interoceptive awareness and empathy as well as in pathological conditions such as autism or alexithymia. After that, we show that the amplitude of empathic brain responses is modulated not only by dispositional factors, but also by contextual factors such as the attitude toward the other person or the appraisal of the situation. We conclude the chapter with suggestions for future research.

Introduction

With the emergence of social neuroscience, researchers have started to investigate the human ability to understand other people's minds, including their emotions (for reviews, see de Vignemont & Singer, 2006; Decety & Jackson, 2004; Leiberg & Anders, 2006; Preston & de Waal, 2002). The ability to share and understand others' emotions is called *empathy*. Empathy constitutes a major component of social intelligence and is essential for successful interactions in our social environment. The inability to understand what people around you are feeling would not only impair your ability to predict their future behavior, as emotions are strong motivators, but also leave you devoid of close relationships. Apart from being important on a personal level, understanding empathy and its neural underpinnings could have fundamental implications on a societal level, as empathy has been identified as a prerequisite for prosocial motivation and behavior.

In this chapter, we will first address the difficult issue of delineating the concept of empathy and provide a definition that allows us to formulate hypotheses that can be tested with neuroscientific methods and that differentiates empathy from other, sometimes synonymously used, concepts. Second, we will present a putative neural account of empathy and review pertinent peripheral and central neurophysiological evidence. Third, we will focus on the insula, a brain structure that has been ascribed a crucial role for empathy, and discuss its empathy-related functions. Fourth, we will review studies of normal populations as well as those with deficient and superior empathic ability, respectively, to demonstrate the importance of studying individual differences to increase our understanding of the neural bases of empathy. Fifth, we will demonstrate that empathic reactions are influenced not only by dispositional, but also by contextual factors. We will conclude with a summary and an outlook on future research.

TANIA SINGER and SUSANNE LEIBERG Laboratory for Social and Neural Systems Research, Institute for Empirical Research in Economics, University of Zurich, Zurich, Switzerland

Definition

In recent years, a growing body of social neuroscience research has been devoted to delineating the neural substrates that underlie our ability to understand the beliefs, intentions, motives, and emotions of other people. In acknowledgment of the fact that our brains do not exist in isolation, but are part of a complex social environment, interactive mind paradigms have been developed to comprehensively study how inferences about others' mind states are drawn (King-Casas et al., 2005; Montague et al., 2002; Sebanz, Knoblich, Prinz, & Wascher, 2006; Singer et al., 2004, 2006; Tomlin et al., 2006). Until recently, separate lines of research have been pursued: one investigating our understanding of others' beliefs and thoughts, also referred to as *mentalizing* or *cognitive perspective taking*, another studying our ability to understand action intentions by simulating others' motor actions, and still another studying how we share others' emotions and sensations, that is, how we empathize with other people. In this chapter, we will describe the neuronal bases of empathy.

Most definitions of the term *empathy* are rather broad. However, to render empathy investigable with neuroscientific methods, a clear definition that distinguishes it from related concepts such as mentalizing, emotional contagion, sympathy, and empathic concern is crucial (see also Batson,

1991, 2008; Decety & Lamm, 2006; Eisenberg, 2000; Eisenberg & Strayer, 1987; Hoffman, 2000; Singer, 2006). De Vignemont and Singer (2006), for example, identified four constitutive factors of empathy: (1) the presence of an affective state in ourselves, (2) isomorphism between our own and another person's affective state, (3) elicitation of our affective state upon observation or imagination of another person's affective state, and (4) knowledge that the other person's affective state is the source of our own affective state. The first aspect differentiates empathy from mentalizing. The term *mentalizing* refers to the drawing of inferences about other people's mental states, including their affective states, but it does not entail emotional involvement. In contrast, empathy (i.e., the sharing of other people's emotions) does entail emotional involvement. Take psychopaths, for example: Psychopaths readily understand others' mental states, including their affective states, but lack the feeling of empathy. Recently, neuroscientific evidence has accumulated demonstrating that the abilities to empathize and to mentalize indeed rely on distinct neural networks (Blair, 2005; Hynes, Baird, & Grafton, 2006; Völlm et al., 2006).

The second aspect differentiates empathy from sympathy (Eisenberg, 2007; Wispé, 1986). Empathy involves sharing another person's emotions, while sympathy entails an emotional response that is congruent with the other person's emotion but not necessarily isomorphic. For example, when someone is angry about something, you might not share the person's anger (empathize), but you might pity him or her (sympathize).

The third aspect of de Vignemont and Singer's definition states that empathy is elicited not only by observing another person's display of emotion, but also by imagining another person's emotion (i.e., putting oneself in someone else's shoes)—for example, when you read a letter from a friend in which he or she describes a sad event or in other situations in which you cannot glean information about another person's affective state from facial, vocal, or other indicators but have to imagine the circumstances that person is facing.

The last aspect of the definition pertains to the distinction between empathy and emotional contagion. The term *emotional contagion* refers to "the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person and, consequently, to converge emotionally" (Hatfield, Cacioppo, & Rapson, 1994) and has been seen as an important precursor to empathy that is already present in infants (Hoffman, 1984). In contrast to empathy, there is no self/other distinction in emotional contagion: The person is not aware that the emotion he or she is having was elicited by the other person. To summarize, empathy shares features with concepts such as mentalizing, sympathy, and emotional contagion, but we argue that only the co-occurrence of the above-mentioned factors makes an affective experience an instance of empathy.

A shared network hypothesis of empathy

How is the isomorphic affective state that we defined as empathy elicited by merely observing or imagining another person's affective state in the absence of any emotional or sensory input to our own body? Theodor Lipps (1903), who introduced the concept of empathy, assumed that we come to know about others' inner states by internally imitating their gestures and actions. This idea of motor mimicry has been revived in experimental psychology and social neuroscience. Inspired by the discovery of mirror neurons, which fire when we perform an action as well as when we merely observe someone else performing that action (Rizzolatti & Craighero, 2004), and by ideomotor theory (Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz, 1987), which assumes that the observation of others performing motor actions automatically activates motor representations in ourselves, Preston and de Waal (2002) proposed a perception-action model of empathy. They suggested that the observation or imagination of a person in an emotional state automatically activates in the observer or imaginer a representation of that emotional state, including its associated autonomic, somatic, and motor responses. This internal simulation then aids one's understanding of the other person's emotions. Supporting evidence has been adduced by peripheral physiological as well as neuroimaging studies.

Peripheral Neurophysiological Evidence for the Shared Network Hypothesis Assuming that activating a representation of an observed emotional state entails the activation of the associated autonomic, somatic, and motor responses renders the shared network hypothesis testable, albeit indirectly, via several means of measurement besides the central neurophysiological one. For example, electromyographic (EMG) studies have shown that looking at different emotional facial expressions evokes corresponding differential changes in facial muscle activity in the observer (Cacioppo, Petty, Losch, & Kim, 1986; Dimberg, 1982; McHugo, Lanzetta, Sullivan, Masters, & Englis, 1985; Vaughan & Lanzetta, 1980; Weyers, Muhlberger, Hefele, & Pauli, 2006). Upon viewing pictures of angry and happy facial expressions, subjects showed increased activity in the corrugator supercilii muscle and the zygomaticus major muscle, respectively (Dimberg, 1982, 1988), even when they were not conscious of which facial expressions they had seen (Dimberg, Thunberg, & Elmehed, 2000). Furthermore, longer onset latencies were observed when subjects viewed

emotional facial expressions and were then able to exhibit the opposite (as compared to the same) facial expression, suggesting that facial motor actions that have been activated concurrently with the neural representation of the depicted emotional state are inhibited (Lee, Dolan, & Critchley, 2008). Indirect evidence for the shared network hypothesis also comes from studies investigating concordance between the observer's and target's autonomic responses. Gottman and Levenson (1985) demonstrated that subjects experienced the same physiological changes when participating in a conversation themselves and when watching a videotape of that conversation later. Furthermore, when subjects watched videotapes of marital interaction and were instructed to rate the feelings of one spouse, a positive relationship was found between the synchrony of the observer's and target's heart rate changes and rating accuracy (Levenson & Ruef, 1992). Harrison, Singer, Rotshtein, Dolan, and Critchley (2006) showed that the subject's pupil size increased linearly with the pupil size of sad facial expressions they viewed.

Central Neurophysiological Evidence for the Shared Network Hypothesis Peripheral neurophysiological studies can provide only indirect evidence for the postulated activation of the cortical representation of an emotional state by the mere observation or imagination of another person in that emotional state. These studies measure the putatively associated motor and autonomic responses. However, the seminal discovery of mirror neurons in the premotor cortex of macaque monkeys provides a possible model for the implementation of the shared network account of empathy on a brain level. These neurons are activated both when the monkey executes certain actions and when the monkey observes the same actions being executed by others (Rizzolatti and Craighero, 2004). While this strong version of the shared network account of understanding others' motor actions (i.e., the same neural population is responsive to information from both visual and motor domains) can be proven in monkey research, it is inherently problematic to test it in humans with single-cell recordings (but see Hutchison, Davis, Lozano, Tasker, & Dostrovsky, 1999).

In neuroimaging research in the domain of action observation, a region is therefore typically considered to be part of a shared network when it is activated during self-generated actions as well as during observation of the same action performed by others. Certainly, overlapping brain activity patterns during action generation and action observation do not necessarily indicate the presence of neurons with mirror-like properties, as the overlap could also result from activation of two distinct subpopulations lying very close to each other (Dinstein, Hasson, Rubin, & Heeger, 2007; Morrison & Downing, 2007). However, the existence of mirror-like networks in human prefrontal and inferior parietal cortex has been ascertained by using functional magnetic resonance imaging (fMRI) (Grezes & Decety, 2001), transcranial magnetic stimulation (TMS) (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995), and magnetoencephalography (MEG) (Hari et al., 1998). These studies revealed that the same neural networks are engaged when subjects execute and observe actions.

Even though neurons with "mirror neuron" properties in the emotional domain have not yet been recorded in the monkey brain, recent fMRI studies on humans in the domain of emotions and empathy suggest that neural networks with mirror-like properties are not restricted to the motor domain or confined to the prefrontal cortex but extend to other brain areas such as somatosensory and insular cortices. Thus evidence is accumulating for the existence of shared neural networks for facial expressions, sensations, and emotions, which enable one to feel—by merely perceiving or imagining another person's sensations or emotions—what the other is feeling.

First evidence for overlapping neural activation during one's own experience and the observation of another person's experience of emotion comes from a study by Wicker and colleagues (2003). Experiencing disgust oneself upon smelling disgusting odors and observing someone else's odor-elicited expression of disgust evoked similar neural responses in the anterior insula (AI) and the anterior cingulate cortex (ACC). A recent study revealed similar overlapping AI activation when subjects drank pleasant and unpleasant drinks and when they saw someone else drinking pleasant and unpleasant drinks and producing the corresponding facial expressions (Jabbi, Swart, & Keysers, 2007).

Most studies on shared networks in the affective domain have been conducted on empathy for pain (Avenanti, Bueti, Galati, & Aglioti, 2005; Avenanti, Paluello, Bufalari, & Aglioti, 2006; Botvinick et al., 2005; Bufalari, Aprile, Avenanti, Di, & Aglioti, 2007; Cheng et al., 2007; Gu & Han, 2007; Jackson, Brunet, Meltzoff, & Decety, 2006; Jackson, Meltzoff, & Decety, 2005; Lamm, Nusbaum, Meltzoff, & Decety, 2007; Lamm, Batson, & Decety, 2007; Lamm & Decety, 2008; Moriguchi et al., 2007; Morrison, Lloyd, di Pellegrino, & Roberts, 2004; Morrison & Downing, 2007; Morrison, Peelen, & Downing, 2007; Saarela et al., 2007; Singer et al., 2004, 2006; Valeriani et al., 2008). For example, Singer and colleagues (2004) recruited couples and measured empathy in vivo by assessing brain activity in the female partner while painful stimulation was applied either to her own or to her partner's right hand via electrodes attached to the back of the hand. The male partner was seated next to the MRI scanner, and a mirror system allowed the female partner to see her own as well as her partner's hand lying on a tilted board in front of her. Differently

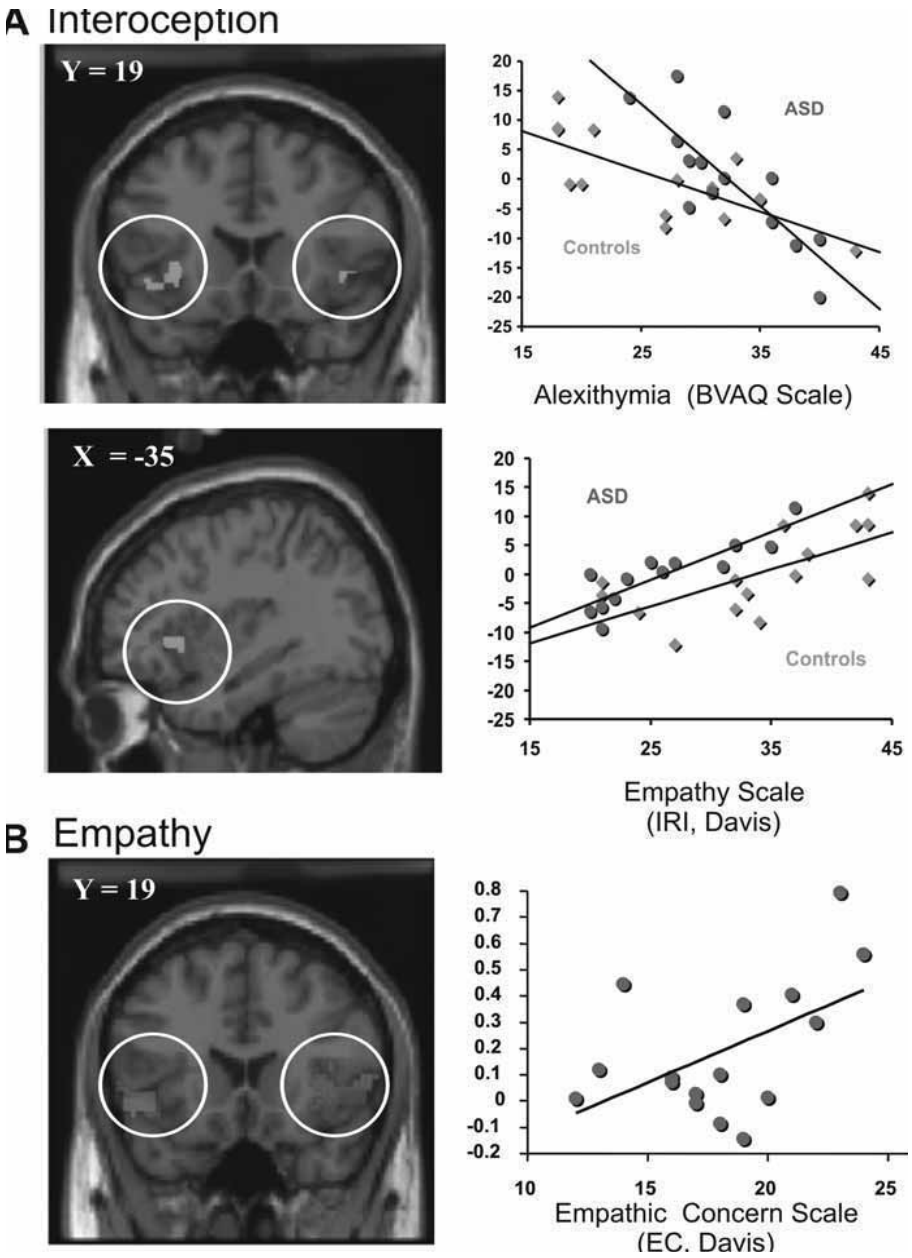


FIGURE 67.1 (*A*) Bilateral AI activation during interoception about one's own feelings correlates with the degree of alexithymia (measured with the Bermond-Vorst Alexithymia Questionnaire by Vorst and Bermond, 2001; BVAQ-B) and empathy (measured with the Interpersonal Reactivity Index by Davis, 1980; IRI) in healthy controls and in subjects with Asperger syndrome (adapted from Silani et al., 2008). (*B*) Overlapping brain activity in bilateral AI in females experiencing pain themselves and empathizing with their partner (red; adapted from Singer et al., 2004) or an unfamiliar but likable person (pink; adapted from Singer et al., 2006) experiencing pain. Individual differences in self-reported empathy (measured with the IRI) covary with activation strengths in AI during empathizing. (From Singer et al., 2004.) (See color plate 81.)

colored flashes of light on a screen pointed to either the male or the female partner's hand, indicating which of them would receive painful stimulation and which would receive nonpainful stimulation. The results suggested that parts of the "pain matrix"—bilateral AI, the rostral ACC, brain stem, and cerebellum—were activated when subjects experienced pain themselves as well as when they saw a signal indicating that their loved one had experienced pain. These areas are involved in the processing of the affective component of pain, that is, how unpleasant the subjectively felt pain is. Thus both the experience of pain to oneself and the knowledge that a beloved partner is experiencing pain activates the same affective pain circuits, suggesting that if a beloved partner suffers pain, our brains also make us suffer from this pain (see figure 67.1). This pattern of shared activation in AI and ACC when one is experiencing pain and

observing others in pain has now been replicated by other studies (Morrison et al., 2004; Singer et al., 2006) and is also consistently observed in studies in which subjects did not receive pain themselves but merely watched videos showing body parts in potentially painful situations (Gu & Han, 2007; Jackson et al., 2006; Jackson, Meltzoff, & Decety, 2005), painful facial expressions (Lamm, Batson, & Decety, 2007; Saarela et al., 2007), and hands being pricked by needles (Cheng et al., 2007; Lamm et al., 2007; Lamm & Decety, 2008; Morrison et al., 2004, 2007).

The finding that these empathic brain responses in the AI and the ACC correlate positively with subjects' trait empathy (Singer et al., 2004), as measured with empathy questionnaires and with unpleasantness ratings of the empathy-eliciting painful stimuli (Jackson, Meltzoff, & Decety, 2005; Lamm et al., 2007; Saarela et al., 2007), has given additional support to the claim that shared networks in the AI and the ACC underlie our ability to empathize with others' suffering (see also figure 67.1).

While fMRI studies probing empathy for pain reliably find activation of brain areas belonging to the affective part of the pain matrix, TMS (Avenanti et al., 2005, 2006), and electroencephalography studies (Bufalari et al., 2007; Valeriani et al., 2008) suggest the additional involvement of sensorimotor parts of the pain matrix in empathy for pain. These studies found reductions in motor- and laser-evoked potentials and increases in somatosensory-evoked potentials when subjects viewed needles penetrating hands. Notably, those changes in evoked potentials correlated with subjective ratings of the sensory but not the affective qualities of the pain and originated in primary and secondary somatosensory cortex. In line with an fMRI study demonstrating the coactivation of the secondary somatosensory cortex during the experience and observation of touch (Keysers et al., 2004), these findings indicate that sensorimotor aspects of pain are also mapped onto the observer's brain.

Insular involvement in interoception, emotion, and empathy

The above-mentioned studies investigating our ability to share others' feelings have ascribed a crucial role to the insula. The question arises as to what general function the insula subserves that is crucial for experiencing our own and others' emotions. We turn to early models of emotions to answer this question. According to the James-Lange theory of emotion, for example, the representation of internal bodily states is a necessary condition for emotional experience to occur. We are in love because our heart beats faster; we are sad because our eyes well up with tears. In recent years, these models have been rediscovered and extended by neuroscientists. Thus, Damasio (1994) formulated the somatic marker hypothesis, which postulates that body signals are a necessary base of a hierarchy of conscious and emotional experiences as well as a crucial influence on motivational behavior especially when we are making risky decisions. More specifically, bodily arousal responses called *somatic markers* are generated in response to the risk of a negative outcome and fed back to unconsciously guide adaptive behavior in the form of "gut feelings" and "hunches."

On the basis of neuroanatomical studies on monkeys Craig (2002, 2003) has proposed that the insula, specifically the AI, underlies the representation and awareness of internal bodily states and thus interoception (i.e., one's sense of the physiological state of one's body). Craig (2002, 2003) suggested that brain afferents convey interoceptive information from the entire body to midbrain reticular nuclei that project to the ACC and posterior dorsal insula via the mediodorsal and ventromedial thalamic nucleus, respectively. The representation of the body's physiological condition in the posterior insula is remapped to the AI, where it is accessible to consciousness. Thus the concomitant projection of stimulus-induced changes in internal bodily states to the ACC and the insula allows for simultaneous awareness of visceral states and the generation of affective behavioral motivation. As the insula is reciprocally connected to multiple brain regions—among them the amygdala, the nucleus accumbens (Reynolds & Zahm, 2005), and the orbitofrontal cortex—it is ideally positioned to integrate contextual information with the bodily effects evoked by this context to a second-order conscious emotional experience. Furthermore, the higher-order representation can have a direct impact on autonomic activity via the insula's bidirectional connections to brain stem and hypothalamic nuclei (Barbas, Saha, Rempel-Clower, & Ghashghaei, 2003).

The notion that AI activity represents information about internal bodily states has been corroborated by several human fMRI studies by Critchley and colleagues. They have shown that activity in the right AI correlated with peripheral physiological indicators of arousal such as skin conductance (Critchley, Elliott, Mathias, & Dolan, 2000; Nagai, Critchley, Featherstone, Trimble, & Dolan, 2004) and cardiovascular responses (Critchley, Corfeld, et al., 2000; Critchley et al., 2005). Moreover, they reported reduced gray matter volume in the insula for patients with pure autonomic failure (PAF), a condition that entails the inability to generate autonomic arousal due to confined peripheral denervation of the autonomic nervous system (Critchley et al., 2003). In an fMRI fear-conditioning paradigm (Critchley, Mathias, & Dolan, 2002), individuals with PAF and controls evinced the typical reaction-time evidence that fear conditioning had taken place. However, subjects with PAF lacked the corresponding autonomic reactions and demonstrated reduced activity in the right middle insula. The role of the AI for interoceptive awareness was studied by introducing a second factor to the experiment, namely,

the awareness of the conditioned stimulus. Remarkably, in accordance with the hypothesis that the AI mediates the conscious experience of internal bodily states, the amygdala was activated by consciously and unconsciously perceived conditioned stimuli, whereas the AI was solely sensitive to consciously perceived conditioned stimuli. Fortifying evidence for the role of the insula in interoceptive awareness, it was demonstrated that the activity and size of the right AI were positively related to the degree to which participants were aware of their own heartbeat (Critchley, Wiens, Rotshtein, Ohman, & Dolan, 2004). Notably, the degree of interoceptive awareness and activity of the right AI were associated with self-reported negative emotional experience. This is in line with a study revealing increasing AI activity with increasing self-reported negative valence of emotional pictures (Anders, Lotze, Erb, Grodd, & Birbaumer, 2004).

The above-mentioned findings from studies addressing empathic responses suggest that the same brain structures (AI and ACC) that are crucial for representing our own affective states also play an important role in sharing the affective states of others. On the basis of this observation, Singer and colleagues (2004) extended an interoceptive model of emotions to the domain of empathy and suggested that cortical rerepresentations of internal bodily states in the AI might serve two functions. First, they might allow us to form subjective representations of our own affective states. These higher-level rerepresentations, however, allow us not only to understand our own affective states when emotional stimuli are present, but also to form forward models that let us predict the bodily effects of anticipated emotional stimuli to our own bodies. Furthermore, these representations may serve as the visceral correlate of a prospective empathic simulation of how something might feel for others. This would then help us to understand the emotional significance of a particular stimulus and its likely consequences.

Interindividual differences in healthy and clinical populations

So far, we have accrued evidence for the hypothesis that being in an affective state and observing another person in that affective state recruit overlapping brain regions and, most important, the insula. To strengthen the claim that a particular brain region does indeed play an essential role for a specific ability, the analysis of interindividual differences in this ability has proven to be a useful approach. Correlating questionnaire measures of empathy—for example, the Interpersonal Reactivity Index (IRI) (Davis, 1983) and the Balanced Emotional Empathy Scale (BEES) (Mehrabian & Epstein, 1972)—with activation strengths or structural indices as well as studying people with deficient or superior empathic abilities can yield additional indications for a functional role of the insula and other brain regions in empathy.

In this line, EMG studies have shown more pronounced mimicking behavior in subjects with high compared to low dispositional empathy when they passively viewed emotional facial expressions (Dimberg, Andreasson, & Thunberg, 2005; Sonnby-Borgström, 2002). Subjects who scored high on dispositional empathy reported greater emotional responding and showed higher SCR than low-empathy subjects did when viewing an empathy-evoking film (Eisenberg et al., 1991). This confirmed earlier findings showing that high-empathy women exhibited larger SCR and more congruent facial expressions than low-empathy women did when watching videotapes of smiling, calm, and crying babies (Wiesenfeld, Whitman, & Malatesta, 1984).

As was mentioned above, analyses of empathic brain responses obtained while subjects were observing other people suffering—be it their loved ones or people they liked (Singer et al., 2004, 2006)—have revealed individual differences in activity in empathy-related pain-sensitive areas (ACC and AI) and covariations of these differences with interindividual differences in IRI and BEES scores (see figure 67.1). The higher subjects scored on these questionnaires, the higher their empathic brain activation in ACC and AI. Interestingly, Jabbi and colleagues (2007) observed similar correlations between IRI subscales and empathic brain responses in the AI in subjects who had observed others tasting pleasant or unpleasant drinks associated with facial expressions of joy or disgust, respectively. Empathic brain responses correlated positively not only with trait measures of empathy, but also with unpleasantness ratings that subjects gave after each trial of an empathy-inducing condition (Jackson, Meltzoff, & Decety, 2005; Lamm et al., 2007; Saarela et al., 2007).

Another way to investigate individual differences in empathy is to study experts in empathizing. Thus, for example, functional and structural changes in empathy-related brain areas have been found in people with a high degree of empathy stemming from long-term compassion meditation. In a recent study based on 15 long-term practitioners, Lutz, Brefczynski-Lewis, Johnstone, and Davidson (2008) demonstrated that during compassion meditation, insula activation by emotional as compared to neutral sounds was stronger in long-term meditators than in novice meditators. This illustrates the malleability of the neural systems underlying empathy and the role of the insula in affect sharing. Similarly, long-term meditators had significantly thicker cortex (Lazar et al., 2005) and greater gray matter volume in the AI (Hölzel et al., 2008).

In contrast to these studies on superior empathic ability, other studies have investigated the deficient empathic ability that is presumed to exist in several psychiatric disorders such as psychopathy and autism (Decety & Moriguchi, 2007). The study of these deficits not only is pertinent for the development of effective therapeutic strategies to treat these

conditions, but also informs us about the mechanisms underlying empathy in the healthy population.

Psychopathy is characterized by a lack of empathy, lack of guilt, and poor behavioral control, a precursor for aggressive behavior (Hare, 1991). Interestingly, people with psychopathy are known to show normal mentalizing ability but seem to lack the ability to empathize with others. Behavioral studies on psychopaths have found selective emotional dysfunction, such as impairment in aversive conditioning, autonomic response to threat, augmentation of the startle reflex to visual threat primes, impaired processing, and altered autonomic response to sad and fearful facial and vocal expressions (Blair, Colledge, Murray, & Mitchell, 2001; Flor, Birbaumer, Hermann, Ziegler, & Patrick, 2002; Levenston, Patrick, Bradley, & Lang, 2000). These deficits have been ascribed to amygdala dysfunction. Indeed, psychopathy is related to reduced gray matter volume in the amygdala (Tiihonen et al., 2000) and reduced amygdala activity during emotional memory (Kiehl et al., 2001) and aversive conditioning (Birbaumer et al., 2005; Veit et al., 2002). Hypoactivation during aversive conditioning in psychopaths has also been observed in the AI (Birbaumer et al., 2005). Importantly, Sterzer, Stadler, Poustka, and Kleinschmidt (2007) found reduced gray matter volume in the amygdala and AI in adolescents with conduct disorder as compared to normal controls. The gray matter volume in bilateral AI but not the amygdala correlated negatively with the empathy scores and observed aggressive behavior, underscoring the important role of insular cortex for the representation of affective states and subsequent appropriate empathic behavior. In a similar vein, boys with disruptive behavior disorder have been found to show lower scores on an empathy questionnaire and significantly less corrugator supercilii activity to angry expressions than did age-matched healthy controls, supporting the notion that deviant social behavior might be a result of diminished empathic responding (de Wied, van Boxtel, Zaalberg, Goudena, & Matthys, 2006). Even though these findings suggest that individuals with psychopathy have empathy deficits but normal mentalizing ability, future neuroimaging studies involving explicit empathy and mentalizing tasks with the same subjects will have to be performed to support this assumption.

Recent studies suggest that, in contrast to psychopathy, autism is associated with a lack of mentalizing ability (for a review, see Frith, 2001) but not necessarily a lack of empathy. Autism is a developmental disorder characterized by deficits in social abilities as well as in language skills (Kanner, 1943). Neuroimaging studies have revealed reduced activation in mentalizing-related brain areas such as the medial prefrontal cortex, the superior temporal sulcus, the temporoparietal junction, and the temporal poles when autistic subjects make inferences about others' mental states (Castelli, Frith, Happe, & Frith, 2002; Happe et al., 1996). The question concerning whether all people with autism have empathy deficits was the subject of a recent discussion (Silani et al., 2008). The inconclusiveness concerning this question might arise from the fact that so far, studies investigating empathy in individuals with autism have not differentiated between subjects with and without alexithymia. Alexithymia, which is present in 50% of the high-functioning patients with autism or Asperger syndrome (Hill, Berthoz, & Frith, 2004), is a condition marked by impaired identification and description of feelings and differentiation of feelings from bodily signals. If we turn to the shared network hypothesis of empathy, which states that representations of one's own emotional states are obligatory for the ability to share others' emotions, an interesting prediction arises: Deficits in understanding one's own emotions, as present in alexithymia, should lead to empathy deficits and be correlated with reduced insula activation. Furthermore, might the prevalence of alexithymia in autism explain the sometimes found empathy deficits in autistic subjects? Silani and colleagues (2008) tested the assumption that subjects with Asperger syndrome and alexithymia, but not subjects with Asperger syndrome but not alexithymia, would evince reduced insula activation when introspecting about their own emotions. As depicted in figure 67.2, this study revealed that when subjects were to judge their own feelings while viewing emotional pictures, the degree of alexithymia but not autism was associated with less activation in the AI. Interestingly, individual differences in the degree of alexithymia correlated highly negatively with individual differences in trait empathy, and levels of both alexithymia and empathy were predictive of AI activation during introspection. These findings are in line with the prediction that deficits in understanding one's own emotions result in empathy deficits and that both should be correlated with lesser activation of the AI. Furthermore, autism per se does not seem to be associated with deficient empathy, but the comorbidity with alexithymia does. This notion awaits further support by studies that are trying to dissociate empathizing from mentalizing abilities using the appropriate paradigms.

The modulation of empathy

The shared network account of empathy and related empirical work presented above might suggest that empathizing proceeds in an automatic manner, albeit modulated by dispositional empathy. Every time we observe or imagine the affective state of another person, we unconsciously share it. However, in line with social psychological research (Batson, Lishner, Cook, & Sawyer, 2005; Zaki, Bolger, & Ochsner, 2008), social neuroscience has accumulated evidence that empathic responses not only underlie the influence of dispositional factors, but also might be modulated by contextual and stimulus-inherent factors and appraisal processes.

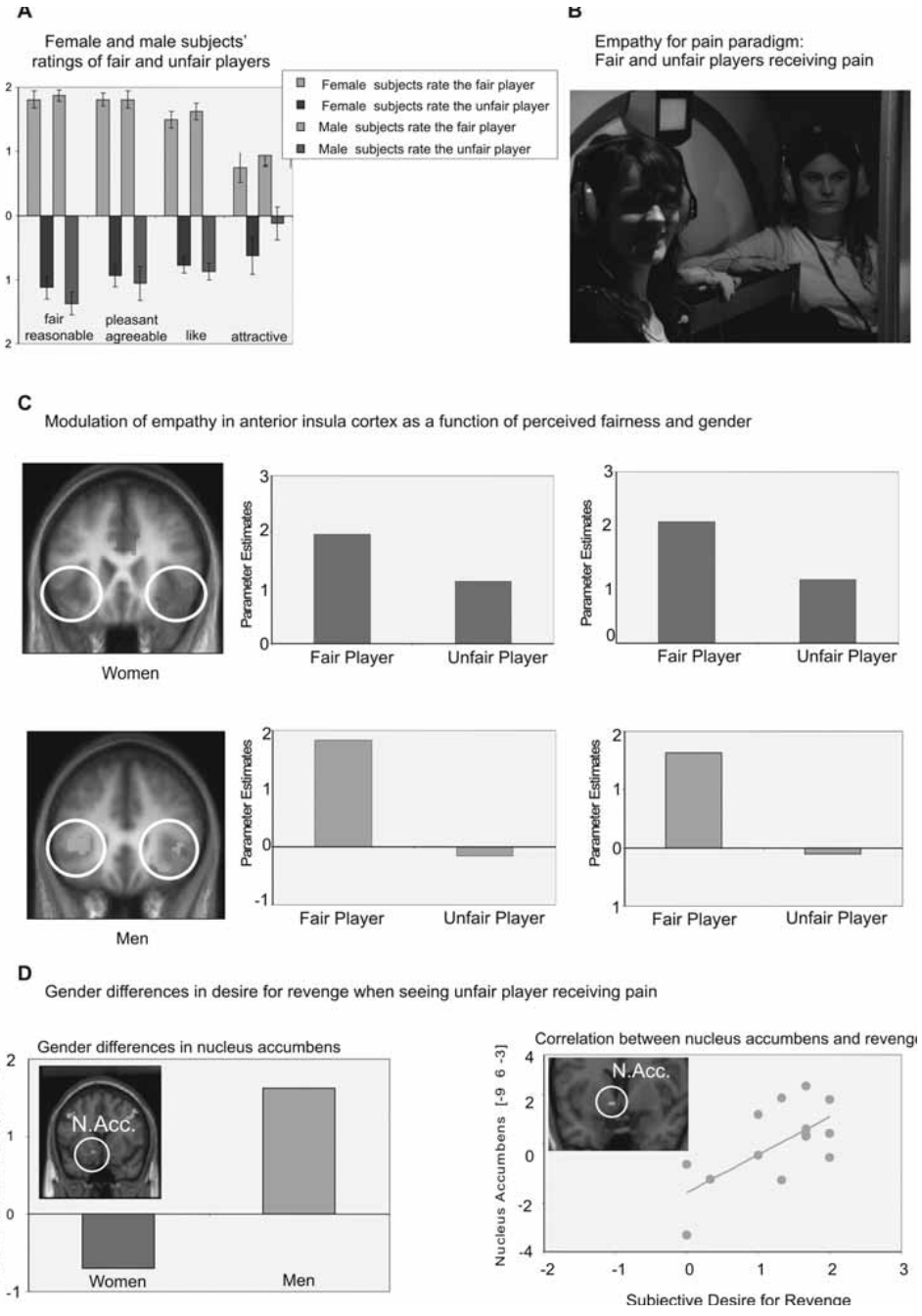


FIGURE 67.2 Modulation of empathic brain responses by perceived fairness (adapted from Singer et al., 2006). (*A*) Female (pink) and male (bule) subject's postscan ratings of perceived fairness, agreeableness, likeability, and attractiveness of the two confederates, who had always played fairly and unfairly, respectively, in a preceding monetary economic trust game. (*B*) Setup of the empathy-for-pain paradigm with the subject lying in the scanner and one fair and one unfair player (both confederates) sitting on either side of the scanner. Electrodes, which were attached to one of their hands, delivered painful or nonpainful stimulation as previously indicated by flashes on a screen in front of them. (*C*) Empathic brain activation in bilateral frontoinsular cortex when males (blue) and females (pink) perceived either the fair or the unfair player suffering pain. Although both men and women showed empathic brain activity in frontoinsular cortex when they perceived a fair player in pain, only women did so when they perceived an unfair player in pain. (*D*) Enhanced activation in nucleus accumbens in men, but not in women, when they perceived unfair as compared to fair players suffering pain. The strength of this activation correlated positively with men's , but not women's, degree of subjectively expressed desire for revenge. (See color plate 82.)

Although facial mimicry has been considered automatic and hard-wired in that it is elicited even without subjects' awareness of the eliciting stimuli, it has been shown that experimentally induced attitudes toward computer avatars modulated the imitative EMG responses (Likowski, Mühlberger, Seibt, Pauli, & Weyers, 2008). A positive attitude toward the avatars evoked facial EMG responses that were congruent with their emotional expressions, whereas a negative attitude resulted in reduced or incongruent facial EMG responses. Furthermore, facial EMG responses have also been shown to be influenced by social group membership (Bourgeois & Hess, 2008; negative emotions were mimicked only when shown by in-group members), one's own affective state (Moody, McIntosh, Mann, & Weisser, 2007), and expectations (Lanzetta & Englis, 1989). Lanzetta and Englis (1989) found a striking influence of context on facial imitation during social interaction. Depending on whether subjects believed that they were cooperating or competing with a co-player in an investment game, their EMG responses were either congruent or incongruent with the co-player's facial expressions. Notably, the only difference between the cooperative and competitive setting was the subject's expectation. The behavior of the co-player as well as the gains and losses were fixed by a computer.

In a similar vein, but using fMRI, Singer and colleagues (2006) demonstrated the modulation of empathic brain responses by the perceived fairness of the other person. In this study, subjects first played a sequential prisoner's dilemma game with two confederates. One confederate played fairly; the other played unfairly. Subsequently, the three participated in a modified version of the empathy-for-pain paradigm by Singer and colleagues (2004) introduced above (see figure 67.2*B*). This time, subject's empathic brain responses were measured while either the subject or one of the confederates was receiving painful stimulation. When the fair player was in pain, empathy-related activation was observed in the subject's AI and the ACC (the latter only in women), again overlapping with activity evoked by the subject experiencing pain himself or herself. However, when observing the unfair player in pain, male, but not female, subjects showed an absence of such empathic activity (see figure 67.2*C*). Instead, men showed increased activation in the reward-related nucleus accumbens, which correlated positively with their desire for revenge as assessed by questionnaires after the scanning session (see figure 67.2*D*). This suggests that at least in men, empathic responding is modulated by perceived personality characteristics of the other person.

Other factors that have been identified by fMRI studies as contributing to intraindividual variation in empathic brain responses are the attention allocated to the stimulus, stimulus reality (Gu & Han, 2007), the subject's appraisal of whether the reason the other person is suffering is justified (Lamm, Batson, & Decety, 2007), and the subject's prior experience with the situation (Cheng et al., 2007). Measuring evoked potentials, Aglioti's group demonstrated a modulatory role of stimulus intensity in that needles penetrating a hand yielded a reduction in motor excitability, whereas needles merely giving a hand a pinprick did not (Avenanti et al., 2006). Furthermore, they showed an influence of being in pain oneself on the brain responses evoked by seeing someone else in pain (Valeriani et al., 2008). Strikingly, not the perceived pain of the model, but the subject's own pain, was predictive of a brain response reduction, which has been associated with empathy. This suggests that people represent others' sensations and emotions more in line with their own feelings than with the other person's feelings.

Neuroscientific research has adduced considerable evidence for the modulation of empathic responses. This brings us to the question: Does this modulation occur before or during an empathic response? De Vignemont and Singer (2006) proposed two possible routes for the influence of appraisal processes. Either an empathic response is automatically initiated by an emotional cue and later modulated by contextual information or the emotional cue is first evaluated in its context and, depending on the outcome of the evaluation, either no empathic response or an enhanced or reduced empathic response is elicited. In the latter model, no automatic initiation of empathic responses occurs. The temporal resolution of fMRI is too coarse to distinguish between these two routes. Hence, electromagnetic studies need to be undertaken to address the timing of modulatory input.

Summary and conclusion

In recent years, major advances have been made in social neuroscience to study empathy in an ecologically valid way and to delineate its causes; underlying neural substrates; and autonomic, peripheral physiological, and behavioral consequences. We share and understand other people's affective states by means of shared affective neural networks that are activated when we experience an affective state as well as when we observe someone else experiencing an affective state. Thus an internal simulation of our own emotions is triggered by observing others' emotions, which provides the basis for understanding others' affective states. It remains to be seen whether the activation of a representation of another person's emotional state is necessary and sufficient for correctly understanding the other person's emotion. So far, the results concerning the relationship between activated representations, subjective feelings, and empathic accuracy are equivocal. There is initial evidence showing that the representation evoked by observing another person's affective state is biased by one's own affective state—a tendency called *egocentricity bias*.

Apart from one's own affective state, other factors have also been shown to modulate empathic responses. Hence, although shared representations might be evoked fairly automatically, they are nevertheless amenable to modulatory input. Future research needs to determine further modulatory factors and when they take effect. Are empathic brain responses modulated from the beginning, or are they influenced at a later point in time? MEG or EEG, with their excellent temporal resolution, or TMS might provide conclusive evidence here.

Neuroimaging studies of empathy have thus far exclusively investigated how we share the *negative* affective states of others, such as pain or disgust. One notable exception is the study by Jabbi and colleagues (2007), which addressed empathy for positive and negative emotions evoked by gustatory stimuli. Interestingly, as in empathy for pain or disgust, the AI showed overlapping activity for experiencing a positive affective state and observing someone else experiencing a positive affective state. Does the insula provide a general emotion-independent neural substrate of empathy, or do distinct neural populations in the insula code for negative and positive affective states? In this line, Craig (2005) proposed that left AI activation is predominantly associated with parasympathetic activity and thus is more closely related to positive affect, approach, and affiliative emotional behavior. In contrast, right AI may be preferentially coupled with efferent sympathetic activity and thus with arousal, negative affect, withdrawal, and survival-related behaviors. Empirical evidence, however, points to an alternative hypothesis: Negative affective states are mainly represented in the anterior part of the insula, whereas positive affective states are represented in the middle insula (Bartels & Zeki, 2004; Gray, Harrison, Wiens, & Critchley, 2007; Lutz et al., 2008). Jabbi and colleagues (2007) showed a correlation of dispositional empathy with bilateral AI activity during observation of disgusted faces and with left AI during observation of happy faces, thus tentatively supporting Craig's hypothesis. However, the search volume in this study was confined to the anterior part of the insula. Hence the question concerning which parts of the insula represent negative and positive affective states awaits further research.

The ability to empathize varies from person to person. This variation not only is reflected in self-reports of dispositional empathy, but is also related to differences in facial mimicry and neural activation in empathy-related brain areas such as the AI and the ACC. Two important issues still need to be clarified: First, can empathic ability be enhanced by means of training? Second, how is empathic ability related to prosocial behavior? Training-induced functional and structural brain changes in humans have been shown in a variety of domains, such as spatial navigation (Maguire et al., 2000), motor (Draganski et al., 2004) and musical training (Besson, Schon, Moreno, Santos, & Magne, 2007), and memory (Draganski et al., 2006). However, no study demonstrating the effects of empathy training on cortical plasticity and behavior has ever been published. Studies reporting structural and functional brain changes in long-term compassion meditators suggest that affective neural networks are malleable. Evidence for positive neural and behavioral effects of empathy training could have important implications for therapeutic approaches to clinical conditions involving empathy deficits as well as for general education.

In social psychology, investigation of the link between empathy and prosocial behavior has a long history (Batson, 1998; Eisenberg & Fabes, 1991). It has been suggested that sharing another person's affect can lead either to empathic concern/sympathy, which is an other-oriented response or to personal distress, depending on one's emotion regulation ability and intensity of engagement with the other person. Empathic concern/sympathy increases prosocial action tendencies, whereas personal distress results in self-oriented responses, such as fleeing the scene, to alleviate one's own distress. Disentangling these two concepts and increasing our understanding of the neural bases of prosocial behavior are pertinent tasks for social neuroscience. It is conceivable that both empathic concern and personal distress are represented in the insula, since both entail interoceptive changes. However, the former can be construed as a positive affective response (e.g., related to love) and the latter as a negative affective response. Therefore the above-mentioned investigation of the neural substrates underlying positive affective states might also yield important insight pertaining to this differentiation. The social neuroscientific study of prosocial behavior is hampered by the lack of suitable experimental paradigms that are applicable to neuroscientific methods. Promising approaches might include the use of virtual reality and economic games. The latter have recently been used in neuroeconomic studies investigating the neural mechanisms underlying decision making in economic settings (de Quervain et al., 2004; Harbaugh, Mayr, & Burghart, 2007; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003) and prosocial behavior (Moll et al., 2006). An integration of paradigms derived from the two emerging fields of social neuroscience and neuroeconomics would allow us to further specify the role of different social emotions such as empathy, our sense of fairness, and our desire for revenge in motivating prosocial and antisocial behavior.

ACKNOWLEDGMENTS This work was supported by the University Research Priority Program "Foundation of Human Social Behavior" of the University of Zürich and by the National Centre of Competence in Research of Neuronal Plasticity and Repair.

REFERENCES

Anders, S., Lotze, M., Erb, M., Grodd, W., & Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: A response-related fMRI study. *Hum. Brain Mapping*, *23*, 200–209.

Avenanti, A., Bueti, D., Galati, G., & Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nat. Neurosci.*, *8*, 955–960.

Avenanti, A., Paluello, I. M., Bufalari, I., & Aglioti, S. M. (2006). Stimulus-driven modulation of motor-evoked potentials during observation of others' pain. *NeuroImage*, *32*, 316–324.

Barbas, H., Saha, S., Rempel-Clower, N., & Ghashghaei, T. (2003). Serial pathways from primate prefrontal cortex to autonomic areas may influence emotional expression. *BMC Neurosci.*, *4*, 25.

Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *NeuroImage*, *21*, 1155–1166.

Batson, C. D. (1998). Altruism and prosocial behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 282–316). Boston: McGraw-Hill.

Batson, C. D. (1991). *The altruism question: Toward a social psychological answer.* Hillsdale, NJ: Lawrence Erlbaum.

Batson, C. D. (2008). These things called empathy. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–15). Cambridge, MA: MIT Press.

Batson, C. D., Lishner, D. A., Cook, J., & Sawyer, S. (2005). Similarity and nurturance: Two possible sources of empathy for strangers. *Basic Appl. Soc. Psychology*, *27*, 15–25.

Besson, M., Schon, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.*, *25*, 399–410.

Birbaumer, N., Veit, R., Lotze, M., Erb, M., Hermann, C., Grodd, W., et al. (2005). Deficient fear conditioning in psychopathy: A functional magnetic resonance imaging study. *Arch. Gen. Psychiatry*, *62*, 799–805.

Blair, R. J. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Conscious. Cogn.*, *14*, 698–718.

Blair, R. J., Colledge, E., Murray, L., & Mitchell, D. G. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *J. Abnorm. Child Psychol.*, *29*, 491–498.

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *NeuroImage*, *25*, 312–319.

Bourgeois, P., & Hess, U. (2008). The impact of social context on mimicry. *Biol. Psychol.*, *77*, 343–352.

Bufalari, I., Aprile, T., Avenanti, A., Di, R. F., & Aglioti, S. M. (2007). Empathy for pain and touch in the human somatosensory cortex. *Cereb. Cortex*, *17*, 2553–2561.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *J. Pers. Soc. Psychol.*, *50*, 260–268.

Castelli, F., Frith, C., Happe, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*, 1839–1849.

Cheng, Y., Lin, C. P., Liu, H. L., Hsu, Y. Y., Lim, K. E., Hung, D., et al. (2007). Expertise modulates the perception of pain in others. *Curr. Biol.*, *17*, 1708–1713.

Craig, A. D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nat. Rev. Neurosci.*, *3*, 655–666.

Craig, A. D. (2003). Interoception: The sense of the physiological condition of the body. *Curr. Opin. Neurobiol.*, *13*, 500–505.

Craig, A. D. (2005). Forebrain emotional asymmetry: A neuroanatomical basis? *Trends Cogn. Sci.*, *9*, 566–571.

Critchley, H. D., Corfield, D. R., Chandler, M. P., Mathias, C. J., & Dolan, R. J. (2000). Cerebral correlates of autonomic cardiovascular arousal: A functional neuroimaging investigation in humans. *J. Physiol., 523* (Pt. 1), 259–270.

Critchley, H. D., Elliott, R., Mathias, C. J., & Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study. *J. Neurosci.*, *20*, 3033–3040.

Critchley, H. D., Good, C. D., Ashburner, J., Frackowiak, R. S., Mathias, C. J., & Dolan, R. J. (2003). Changes in cerebral morphology consequent to peripheral autonomic denervation, *NeuroImage*, *18*, 908–916.

Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2002). Fear conditioning in humans: The influence of awareness and autonomic arousal on functional neuroanatomy. *Neuron*, *33*, 653–663.

Critchley, H. D., Rotshtein, P., Nagai, Y., O'Doherty, J., Mathias, C. J., & Dolan, R. J. (2005). Activity in the human brain predicting differential heart rate responses to emotional facial expressions. *NeuroImage*, *24*, 751–762.

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nat. Neurosci.*, *7*, 189–195.

Damasio, A. R. (1994). Descartes' error and the future of human life. *Sci. Am.*, *271*, 144.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *J. Pers. Soc. Psychol.*, *44*, 113–126.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–1258.

de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends Cogn. Sci.*, *10*, 435–441.

de Wied, M., van Boxtel, A., Zaalberg, R., Goudena, P. P., & Matthys, W. (2006). Facial EMG responses to dynamic emotional facial expressions in boys with disruptive behavior disorders. *J. Psychiatr. Res.*, *40*, 112–121.

Decety, J. & Jackson, P. L. (2004). The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.*, *3*, 71–100.

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *Sci. World J.*, *6*, 1146–1163.

Decety, J., & Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: Implications for intervention across different clinical conditions. *Biopsychosoc. Med.*, *1*, 22.

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology*, *19*, 643–647.

Dimberg, U. (1988). Facial electromyography and the experience of emotion. *J. Psychophysiol.*, *2*, 277–282.

Dimberg, U., Andreasson, P., & Thunberg, M. (2005). Empathy and facial reactions to facial expressions. *Psychophysiology*, *42*(Suppl. 1), S50.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.*, *11*, 86–89.

Dinstein, I., Hasson, U., Rubin, N., & Heeger, D. J. (2007). Brain areas selective for both observed and executed movements. *J. Neurophysiol.*, *98*, 1415–1427.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature*, *427*, 311–312.

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H. G., Winkler, J., Buchel, C., & May, A. (2006). Temporal and spatial dynamics of brain structure changes during extensive learning. *J. Neurosci.*, 26: 6314–6317.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annu. Rev. Psychol.*, *51*, 665–697.

Eisenberg, N. (2007). Empathy-related responding and prosocial behaviour. *Novartis. Found. Symp.*, 278, 71–80.

Eisenberg, N., & Fabes, R. A. (1991). Prosocial behavior and empathy: A multimethod, developmental perspective. In M. Clark (Ed.), *Review of personality and social psychology* (pp. 34–61). Newbury Park, CA: Sage.

Eisenberg, N., Fabes, R. A., Schaller, M., Miller, P., Carlo, G., Poulin, R., et al. (1991). Personality and socialization correlates of vicarious emotional responding. *J. Pers. Soc. Psychol.*, *61*, 459–470.

Eisenberg, N., & Strayer, J. A. 1987. *Empathy and its development.* Cambridge, UK: Cambridge University Press.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *J. Neurophysiol.*, *73*, 2608–2611.

Flor, H., Birbaumer, N., Hermann, C., Ziegler, S., & Patrick, C. J. (2002). Aversive Pavlovian conditioning in psychopaths: Peripheral and central correlates. *Psychophysiology*, *39*, 505–518.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron*, *32*, 969–979.

Gottman, J. M., & Levenson, R. W. (1985). A valid measure for obtaining self-report of affect. *J. Consult. Clin. Psychol.*, *53*, 151–160.

Gray, M. A., Harrison, N. A., Wiens, S., & Critchley, H. D. (2007). Modulation of emotional appraisal by false physiological feedback during fMRI. *PLoS.ONE*, *2*, e546.

Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Hum. Brain Mapping*, *12*, 1–19.

Gu, X., & Han, S. (2007). Attention and reality constraints on the neural processes of empathy for pain. *NeuroImage*, *36*, 256–267.

Happe, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., et al. (1996). "Theory of mind" in the brain: Evidence from a PET scan study of Asperger syndrome. *NeuroReport*, *8*, 197–201.

Harbaugh, W. T., Mayr, U., & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, *316*, 1622–1625.

Hare, R. D. (1991). *The Hare psychopathy checklist—Revised.* Toronto: Multi-Health Systems.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proc. Natl. Acad. Sci. USA*, *95*, 15061–15065.

Harrison, N. A., Singer, T., Rotshtein, P., Dolan, R. J., & Critchley, H. D. (2006). Pupillary contagion: Central mechanisms engaged in sadness processing. *Soc. Cogn. Affect. Neurosci.*, *1*, 5–17.

Hatfield, E., Cacioppo, J. T., & Rapson, R. (1994). *Emotional contagion.* New York: Cambridge University Press.

Hill, E., Berthoz, S., & Frith, U. (2004). Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *J. Autism Dev. Disord.*, *34*, 229–235.

Hoffman, M. L. (1984). Interaction of affect and cognition in empathy. In C. Izard, J. Kagan, & R. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 103–131). New York: Cambridge University Press.

Hoffman, M. L. (2000). *Empathy and moral development.* Cambridge, UK: Cambridge University Press.

Hölzel, B. K., Ott, U., Gard, T., Hempel, H., Weygandt, M., Morgen, K., et al. (2008). Investigation of mindfulness meditation practitioners with voxel-based morphometry. *Scandinavica Int. J. Scand. Studies*, *3*, 55–61.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behav. Brain Sci.*, *24*, 849–878.

Hutchison, W. D., Davis, K. D., Lozano, A. M., Tasker, R. R., & Dostrovsky, J. O. (1999). Pain-related neurons in the human cingulate cortex. *Nat. Neurosci.*, *2*, 403–405.

Hynes, C. A., Baird, A. A., & Grafton, S. T. (2006). Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia*, *44*, 374–383.

Jabbi, M., Swart, M., & Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *NeuroImage*, *34*, 1744–1753.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia*, *44*, 752–761.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage*, *24*, 771–779.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, *42*, 335–346.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., et al. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biol. Psychiatry*, *50*, 677–684.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, 308, 78–83.

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *J. Cogn. Neurosci.*, *19*, 42–58.

Lamm, C., & Decety, J. (2008). Is the extrastriate body area (EBA) sensitive to the perception of pain in others? *Cereb. Cortex 18*, 2369–2373.

Lamm, C., Nusbaum, H. C., Meltzoff, A. N., & Decety, J. (2007). What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. *PLoS ONE*, *2*, e1292. doi: 10.1371/journal.pone.0001292.

Lanzetta, J. T., & Englis, B. G. (1989). Expectations of cooperation and competition and their effects on observers' vicarious emotional responses. *J. Pers. Soc. Psychol.*, *56*, 543–554.

Lazar, S. W., Kerr, C. E., Wasserman, R. H., Gray, J. R., Greve, D. N., Treadway, M. T., et al. (2005). Meditation

experience is associated with increased cortical thickness. *NeuroReport*, *16*, 1893–1897.

Lee, T. W., Dolan, R. J., & Critchley, H. D. (2008). Controlling emotional expression: Behavioral and neural correlates of nonimitative emotional responses. *Cereb. Cortex*, *18*, 104–113.

Leiberg, S., & Anders, S. (2006). The multiple facets of empathy: A survey of theory and evidence. *Prog. Brain Res.*, *156*, 419–440.

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *J. Pers. Soc. Psychol.*, *63*, 234–246.

Levenston, G. K., Patrick, C. J., Bradley, M. M., & Lang, P. J. (2000). The psychopath as observer: Emotion and attention in picture processing. *J. Abnorm. Psychol.*, *109*, 373–385.

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., & Weyers, P. (2008). Modulation of facial mimicry by attitudes. *J. Exp. Soc. Psychol.*, *44*, 1065–1072.

Lipps, T. (1903). Einfühlung, innere Nachahmung, und Organempfindungen [Empathy, inner imitation, and sense-feelings]. *Arch. Gesamte Psychol.*, *1*, 185–204.

Lutz, A., Brefczynski-Lewis, J., Johnstone, T., & Davidson, R. J. (2008). Regulation of the neural circuitry of emotion by compassion meditation: Effects of meditative expertise. *PLoS ONE*, *3*, e1897. doi:10.1371/journal.pone.0001897

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proc. Natl. Acad. Sci. USA*, *97*, 4398–4403.

McHugo, G. J., Lanzetta, J. T., Sullivan, D. G., Masters, R. D., & Englis, B. G. (1985). Emotional reactions to a political leader's expressive displays. *J. Pers. Soc. Psychol.*, *49*, 1513–1529.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *J. Pers.*, *40*, 525–543.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. USA*, 103, 15623–15628.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., et al. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage*, *16*, 1159–1164.

Moody, E. J., McIntosh, D. N., Mann, L. J., & Weisser, K. R. (2007). More than mere mimicry? The influence of emotion on rapid facial reactions to faces. *Emotion*, *7*, 447–457.

Moriguchi, Y., Decety, J., Ohnishi, T., Maeda, M., Mori, T., Nemoto, K., et al. (2007). Empathy and judging other's pain: An fMRI study of alexithymia. *Cereb. Cortex*, *17*, 2223–2234.

Morrison, I., & Downing, P. E. (2007). Organization of felt and seen pain responses in anterior cingulate cortex. *NeuroImage*, *37*, 642–651.

Morrison, I., Lloyd, D., di Pellegrino, G., & Roberts, N. (2004). Vicarious responses to pain in anterior cingulate cortex: Is empathy a multisensory issue? *Cogn. Affect. Behav. Neurosci.*, *4*, 270–278.

Morrison, I., Peelen, M. V., & Downing, P. E. (2007). The sight of others' pain modulates motor processing in human cingulate cortex. *Cereb. Cortex*, *17*, 2214–2222.

Nagai, Y., Critchley, H. D., Featherstone, E., Trimble, M. R., & Dolan, R. J. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: A physiological account of a "default mode" of brain function. *NeuroImage*, *22*, 243–251.

Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behav. Brain Sci.*, *25*, 1–20.

Prinz, W. (1987). Ideomotor action. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 47–76). Hillsdale, NJ: Lawrence Erlbaum.

Reynolds, S. M., & Zahm, D. S. (2005). Specificity in the projections of prefrontal and insular cortex to ventral striatopallidum and the extended amygdala. *J. Neurosci.*, *25*, 11757–11767.

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *NeuroReport*, *15*, 2539–2543.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, *27*, 169–192.

Saarela, M. V., Hlushchuk, Y., Williams, A. C., Schurmann, M., Kalso, E., & Hari, R. (2007). The compassionate brain: Humans detect intensity of pain from another's face. *Cereb. Cortex*, *17*, 230–237.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, *300*, 1755–1758.

Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: An ERP study of action planning and control in co-acting individuals. *J. Cogn. Neurosci.*, *18*, 859–870.

Silani, G., Bird, G., Brindley, R., Singer, T., Frith, U., & Frith, C. (2008). Levels of emotion awareness and autism: An fMRI study. *Soc. Neurosci.*, *3*, 97–112.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neurosci. Biobehav. Rev.*, *30*, 855–863.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*, 1157–1162.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*, 466–469.

Sonnby-Borgström, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scand. J. Psychol.*, *43*, 433–443.

Sterzer, P., Stadler, C., Poustka, F., & Kleinschmidt, A. (2007). A structural neural deficit in adolescents with conduct disorder and its association with lack of empathy. *NeuroImage*, *37*, 335–342.

Tiihonen, J., Hodgins, S., Vaurio, O., Laakso, M., Repo, E., Soininen, H., et al. (2000). Amygdaloid volume loss in psychopathy. *Soc. Neurosci. [Abstracts]*, *26*, 2017.

Tomlin, D., Kayali, M. A., King-Casas, B., Anen, C., Camerer, C. F., Quartz, S. R., et al. (2006). Agent-specific responses in the cingulate cortex during economic exchanges. *Science*, *312*, 1047–1050.

Valeriani, M., Betti, V., Le, P. D., De, A. L., Miliucci, R., Restuccia, D., et al. (2008). Seeing the pain of others while being in pain: A laser-evoked potentials study. *NeuroImage*, *40*, 1419–1428.

Vaughan, K. B., & Lanzetta, J. T. (1980). Vicarious instigation and conditioning of facial expressive and autonomic responses to a model's expressive display of pain. *J. Pers. Soc. Psychol.*, *38*, 909–923.

Veit, R., Flor, H., Erb, M., Hermann, C., Lotze, M., Grodd, W., et al. (2002). Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. *Neurosci. Lett.*, *328*, 233–236.

Völlm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., et al. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, *29*, 90–98.

Vorst, H. C. M., & Bermond, B. (2001). Validity and reliability of the Bermond-Vorst alexithymia questionnaire. *Pers. Indiv. Differ.*, *30*, 413–434.

Weyers, P., Muhlberger, A., Hefele, C., & Pauli, P. (2006). Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology*, *43*, 450–453.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*, 655–664.

Wiesenfeld, A. R., Whitman, P. B., & Malatesta, C. Z. (1984). Individual differences among adult women in sensitivity to infants: Evidence in support of an empathy concept. *J. Pers. Soc. Psychol.*, *46*, 118–124.

Wispé, L. (1986). The distinction between sympathy and empathy: To call forth a concept, a word is needed. *J. Pers. Soc. Psychol.*, *50*, 314–321.

Zaki, J., Bolger, N., & Ochsner, K. N. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychol. Sci.*, *19*, 399–404.

68 The Cognitive Neuroscience of Moral Judgment

JOSHUA D. GREENE

ABSTRACT This article reviews recent advances in the cognitive neuroscience of moral judgment. The field began with studies of individuals who exhibit abnormal moral behavior, including neurological patients and psychopaths. Such studies continue to provide valuable insights, particularly concerning the role of emotion in moral decision making. Recent functional neuroimaging studies of normal individuals have identified neural correlates of specific emotional processes relevant to moral judgment. A range of studies using diverse methods support a dual-process theory of moral judgment according to which utilitarian moral judgments (favoring the "greater good" over individual rights) are enabled by controlled cognitive processes, while deontological judgments (favoring individual rights) are driven by intuitive emotional responses. Several recent neuroimaging studies focus on the neural bases of mental state attribution in the context of moral judgment. Finally, research in the field of neuroeconomics has focused on neural processing related to cooperation, trust, and fairness.

The aim of cognitive neuroscience is to understand the mind in physical terms. This endeavor assumes that the mind *can* be understood in physical terms and, insofar as it is successful, validates that assumption. Against this philosophical backdrop, the cognitive neuroscience of moral judgment takes on special significance. Moral judgment is, for many, the quintessential operation of the mind beyond the body, the earthly signature of the soul (Greene, in press). (In many religious traditions, it is, after all, the quality of a soul's moral judgment that determines where it ends up.) Thus the prospect of understanding moral judgment in physical terms is especially alluring or unsettling, depending on your point of view.

In this brief review, I provide a progress report on our attempts to understand how the human brain makes moral judgments. In recent years, we have continued to learn valuable lessons from individuals whose abnormal brains dispose them to abnormal social behavior. We have developed new moral-psychological testing materials and have used them to dissociate and characterize the affective and cognitive processes that shape moral decisions. Finally, the field of neuroeconomics has brought a welcome dose of ecological validity to the study of moral decision making. I discuss each of these developments below. (Important and relevant developments in other fields, such as animal behavior and developmental psychology (de Waal, 2006; Hamlin, Wynn, & Bloom, 2007; Warneken, Hare, Melis, Hanus, & Tomasello, 2007), are beyond the scope of this article.)

JOSHUA D. GREENE Department of Psychology, Harvard University, Cambridge, Massachusetts

Bad brains

In the 1990s, Damasio and colleagues published a series of path-breaking studies of decision making in patients with damage to ventromedial prefrontal cortex (VMPFC), one of the regions that was damaged in the famous case of Phineas Gage (Damasio, 1994; Macmillan, 2000). VMPFC patients were mysterious because their real-life decision making was clearly impaired by their lesions, but their deficits typically evaded detection when standard neurological measures of executive function were used (Saver & Damasio, 1991). Notably, such patients showed no sign of impairment on Kohlberg's (Colby & Kohlberg, 1987) widely used test of moral reasoning (Anderson, Bechara, Damasio, Tranel, & Domasio, 1999). Using a game designed to simulate real-world risky decision making (the Iowa Gambling Task), Bechara, Tranel, Damasio, and Damasio (1996) documented these behavioral deficits and demonstrated, using autonomic measures, that these deficits are emotional. It seems that such patients make poor decisions because they are unable to generate the feelings that guide adaptive decision making in healthy individuals.

A later study targeting moral judgment (Anderson et al., 1999) compared patients with adult-onset VMPFC damage to two patients who had acquired VMPFC damage as young children. While the late-onset patients make poor real-life decisions (e.g., neglecting relatives and friends, involvement in shady business ventures), indicating a deterioration of "moral character" (Damasio, 1994), their behavior tended to harm themselves as much as others. The early-onset patients, however, developed into "sociopathic" adults who, in addition to being irresponsible and prone to risk taking, were duplicitous, aggressive, and strikingly lacking in empathy. Furthermore, these two patients, unlike the late-onset patients, exhibited a childlike "preconventional" pattern of

moral judgment, reasoning about moral issues from an egocentric perspective that focused on reward and punishment. This result suggests a critical role for emotion in moral development. The late-onset patients are prone toward bad decision making, but thanks to a lifetime of emotional experience, they are not truly sociopathic. The early-onset patients, in contrast, lacked the emotional responses necessary to learn the basics of human moral behavior (see also Grattan & Eslinger, 1992).

Studies of psychopaths and other individuals with antisocial personality disorder (APD) underscore the importance of emotion in moral decision making. APD is a catch-all diagnosis for individuals whose behavior is unusually antisocial. Psychopathy, in contrast, is a more specific, somewhat heritable (Blonigen, Hicks, Krueger, Patrick, & Iacono, 2005; Viding, Blair, Moffitt, & Plomin, 2005) disorder whereby individuals exhibit a pathological degree of callousness, lack of empathy or emotional depth, and lack of genuine remorse for their antisocial actions (Hare, 1991). Psychopaths tend to engage in instrumental aggression, while other individuals with APD are characterized by reactive aggression (Berkowitz, 1993; Blair, 2001).

Psychopathy is characterized by profound but selective emotional deficits. Psychopaths exhibit normal electrodermal responses to threat cues (e.g., a picture of shark's open mouth) but reduced responses to distress cues (e.g., a picture of a crying child) (Blair, Jones, Clark, & Smith, 1997). In a particularly revealing study, Blair (1995) demonstrated that psychopaths fail to distinguish between rules that authorities cannot legitimately change ("moral" rules, such as a classroom rule against hitting) and rules that authorities can legitimately change ("conventional" rules, such as a rule prohibiting talking out of turn). According to Blair, psychopaths see all rules as *mere* rules because they lack the emotional responses that lead ordinary people to imbue moral rules with genuine, authority-independent moral legitimacy.

Findings concerning the specific neural bases of psychopathy and APD are varied, implicating a wide range of brain regions, including the orbital frontal cortex (OFC)/VMPFC, insula, anterior cingulate cortex (ACC), posterior cingulate cortex (PCC), amygdala, parahippocampal gyrus, and superior temporal gyrus (Kiehl, 2006; Raine & Yang, 2006; Muller et al., 2008). Blair (2004, 2007) has proposed that psychopathy arises primarily from amygdala dysfunction, which is crucial for stimulus-reinforcement learning (Davis & Whalen, 2001) and thus for normal moral socialization (Oxford, Cavell, & Hughes, 2003). The amygdala exhibits reduced activity in psychopaths both in nonmoral contexts (e.g., in response to emotional words (Kiehl et al., 2001)) and in sociomoral contexts (e.g., during cooperative behavior in a prisoner's dilemma game (Rilling et al., 2007)). Consistent with this view, Yang, Raine, Narr, Lencz, and Toga (2006) found reduced amygdala volume in psychopaths. The VMPFC, which is known to work in concert with the amygdala (Diergaarde, Gerrits, Brouwers, & van Ree, 2005; Schoenbaum & Roesch, 2005), also exhibits many of these effects and appears to play a role in (mis)representing the value of behavioral outcomes in psychopathy (Blair, 2007).

A broader suite of brain regions have been implicated in APD (Raine & Yang, 2006), suggesting, among other things, more general deficits in prefrontal function (Raine et al., 1994). These may be due to structural abnormalities involving reduced prefrontal gray matter (Raine, Lencz, Bihrle, LaCasse, & Colletti, 2000; Yang et al., 2005). There is some evidence implicating abnormal function in dorsolateral prefrontal cortex (DLPFC) in patients with APD (Schneider et al., 2000; Vollm, Richardson, & Stirling, 2004) but not in patients with psychopathy. Given the DLPFC's role in cognitive control (Miller & Cohen, 2001), this is consistent with the notion that psychopaths' aggression results from lack of empathy for others rather than poor impulse control.

Mapping moral emotion

Consistent with research on APD and in keeping with a broader trend in moral psychology (Haidt, 2001), most research using functional imaging to study morality has focused on mapping the "where" and "when" of moral emotion in the brain. Some early studies compared moral and nonmoral stimuli (Moll, Eslinger, & Oliveira-Souza, 2001; Moll, de Oliveira-Souza, Bramati, & Grafman, 2002; Moll, de Oliveira-Souza, Eslinger, et al., 2002) and identified a suite of brain regions that are sensitive to moral stimuli, including the OFC, mPFC, frontal pole, PCC/precuneus, superior temporal sulcus (STS), and temporal pole. This approach, while informative, depends critically on the choice of nonmoral control stimuli and the assumption that the observed results are in some way specific to morality (Greene & Haidt, 2002). More recent functional imaging studies have focused on identifying and functionally characterizing different kinds of moral-emotional processes.

Empathy, Caring, and Harm Greene, Sommerville, Nystrom, Darley, and Cohen (2001) identified a set of brain regions associated with judging actions involving "personal," as compared to "impersonal," harm: mPFC (BA 9/10), the PCC/Precuneus (BA 31), and the posterior superior temporal sulcus (pSTS)/temperoparietal junction (TPJ)/angular gyrus (BA 39). (See figures 68.1 and 68.2 and the section entitled "Dual-Process Morality" below for more details.) A study replicating these results using a larger sample (Greene, Nystrom, Engell, Darley, & Cohen, 2004) identified the same effect in the amygdala, among other regions. The aforementioned regions are implicated in emotional processing (Maddock, 1999; Phan, Wager, Taylor, &

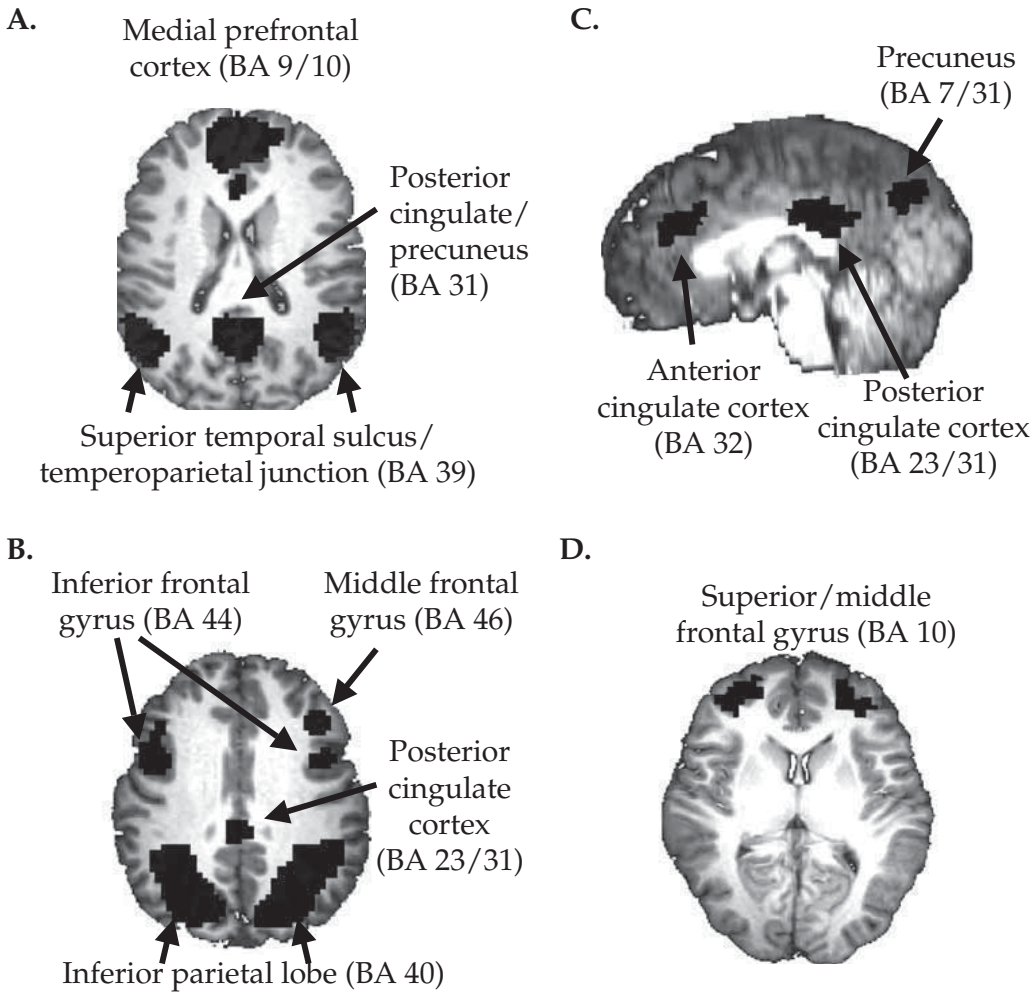


FIGURE 68.1 (*A*) Brain regions exhibiting greater activity in response to personal, as compared to impersonal, moral dilemmas. These regions are part of the "default network" (Gusnard & Raichle, 2001) and are widely activated in many functional imaging studies of moral judgment. (*B*) Brain regions exhibiting greater activity in response to impersonal, as compared to personal, moral dilemmas. (*C*) Brain regions exhibiting greater activity in response to difficult (high-conflict) personal moral dilemmas, as compared to easy (low-conflict) personal moral dilemmas. (*D*) Brain regions exhibiting greater activity in association with utilitarian, as compared to deontological/nonutilitarian moral judgment. (Data are from Greene et al., 2004, Data in panels *A–B* replicate effects originally reported in Greene et al., 2001.)

Liberzon, 2002; Adolphs, 2003) as well as in theory of mind (ToM) (Frith, 2001; Adolphs, 2003; Amodio & Frith, 2006; Young, Cushman, Hauser, & Saxe, 2007). These regions, with the exception of the amygdala, are also part of the "default network" (Gusnard & Raichle, 2001; Fox et al., 2005), a set of brain regions that exhibits relatively high levels of tonic activity and that reliably decreases in activity during outwardly directed tasks. Parts of this network are also implicated in self-referential processing (Akbudak, Shulman, & Raichle, 2001; Kelley et al., 2002), episodic memory, "prospection" (Buckner & Carrol, 2007; Schacter, Addis, & Buckner, 2007), and mind-wandering (Mason et al., 2007). A persistent theme among these processes is the representation of events beyond the observable here and now, such as past, future, and imagined events and mental states. Thus the activity that is observed in this network during the contemplation of dilemmas involving "personal" harm is probably related to the fact that these stimuli involve such nonsensory representations, although this alone does not explain why "personal" dilemmas engage this network more than "impersonal" ones. Consistent with this idea, the functional imaging studies of moral judgment that have most robustly engaged this network involve more complex, text-based, narrative stimuli (Greene et al., 2001, 2004; Schaich Borg et al., 2006, 2008; Robertson et al., 2007; Young et al., 2007; Greene et al., 2009b, 2009c; Schaich Borg, Lieberman, & Kiehl, 2008; Young & Saxe, 2008; Kedia, Berthoz, Wessa, Hilton, & Martinot, 2008).

Several studies have focused on neural responses to different types of harm. Luo and colleagues (2006) found that

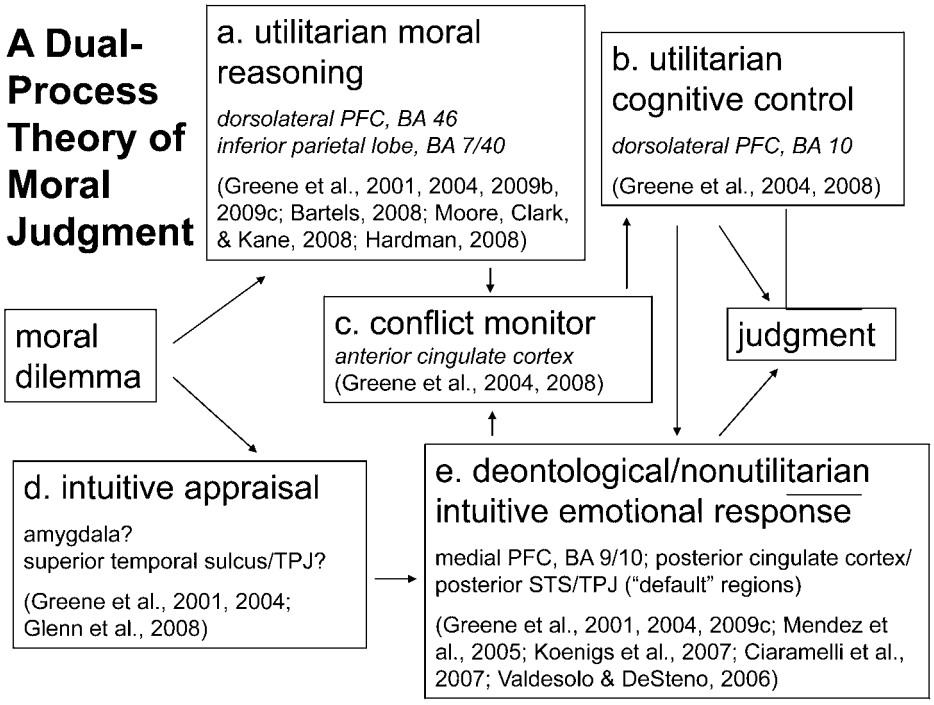


FIGURE 68.2 Diagram of Greene and colleagues' dual-process theory of moral judgment. Boxes a–e describe psychological processes, candidate neural substrates, and references to supporting evidence.

the right amygdala and left VMPFC are sensitive to the intensity of harm displayed in pictures, and Heekeren and colleagues (2005) found that the amygdala exhibits increased activity in response to narratives involving bodily harm. An earlier study (Heekeren et al., 2003) found no effects in the amygdala using stimuli devoid of violence. Finally, individuals with high psychopathy scores exhibited decreased amygdala activity during the contemplation of moral dilemmas involving "personal" harm (Glenn, Raine, & Schug, 2008). Thus the evidence from functional imaging suggests that the amygdala plays an important role in triggering emotional responses to physically harmful actions.

SPECIFIC EMOTIONS Several studies of moral judgment have focused on specific moral emotions, including moral disgust (Rozin, Lowery, & Ebert, 1994; Rozin, Lowery, Imada, & Haidt, 1999; Wheatley & Haidt, 2005). Moll and colleagues (2005) identified a number of brain regions that are sensitive to core/pathogen disgust in moral contexts. A more recent study (Schaich Borg et al., 2008) compared disgust in response to incestuous acts to pathogen disgust and moral transgressions of a nonsexual nature. Stimuli describing nonsexual moral transgressions (e.g., lying, cheating, stealing), as compared to pathogen-disgust stimuli, elicited increased activity in the familiar mPFC/PCC/TPJ network but also in the frontal pole/DLPFC and the ACC. Incest descriptions, as compared to descriptions of nonsexual moral transgressions, elicited increased activity in the mPFC/PCC/TPJ network and other regions, including the inferior frontal gyrus, the left insula, the ventral and dorsal ACC, the left amygdala, and the caudate nucleus. Perhaps surprisingly (Phillips et al., 1997; Calder, Lawrence, & Young, 2001), the insula was preferentially activated only in the incest condition.

Other studies have focused on social emotions such as guilt, embarrassment, shame, pride, empathy, and anger. Robertson and colleagues (2007) found that stimuli associated with care-based morality, as compared to justice-based morality, elicited greater activity in the mPFC and OFC, while the reverse effect was observed in the intraparietal sulcus. In a notably clever study, Beer, Heerey, Keltner, Scabini, and Knight (2003) observed that patients with OFC damage exhibited an inappropriate lack of embarrassment when given an opportunity to disclose personal information, inappropriate embarrassment when overpraised for an unremarkable performance on a simple task, and inappropriate pride and lack of embarrassment when describing the nickname they had invented for the experimenter. Berthoz, Grezes, Armony, Passingham, and Dolan (2006) found that the amygdala is especially responsive to the evaluation of intentional transgressions committed (hypothetically) by oneself (guilt), while Kedia and colleagues (2008) observed increased activity in the mPFC, precuneus, and TPJ for evaluations of transgressions involving others (guilt, anger, compassion). These researchers also observed increased activity in the amygdala, ACC, and basal ganglia for trans-

gressions in which both the self and another are involved (guilt, anger) (see also Shin et al., 2000; Berthoz, Armony, Blair, & Dolan, 2002; Takahashi et al., 2004).

Moral Emotion in Context Other studies have examined the contextual modulation of moral emotion. King, Blair, Mitchell, Dolan, and Burgess (2006) used a custom-designed video game to examine violent versus compassionate behavior in situations in which the behavior is either appropriate (harming an aggressive enemy, helping a distressed innocent person) or inappropriate (helping an aggressive enemy, harming a distressed innocent person). They found that appropriate behavior (whether violent or compassionate) was associated with increased activity in the amygdala and VMPFC. Harenski and Hamann (2006) found that increased mPFC activity and decreased amygdala activity were specifically associated with moral emotion-regulation. Finally, Finger, Marsh, Kamel, Mitchell, and Blair (2006) found that stimuli describing moral/social transgressions committed in the presence of an audience elicited increased activity in the amygdala, underscoring the importance of the amygdala in the social regulation of transgressive behavior (Blair, 2007).

Dual-process morality

The research described above emphasizes the role of emotion in moral judgment (Haidt, 2001), while traditional theories of moral development emphasize the role of controlled cognition (Kohlberg, 1969; Turiel, 2006). My collaborators and I have developed a dual-process theory (Kohlberg, 1969; Posner & Snyder, 1975; Chaiken & Trope, 1999; Lieberman, Gaunt, Gilbert, & Trope, 2002; Kahneman, 2003) of moral judgment that synthesizes these perspectives (see Figure 68.2). According to this theory, both intuitive emotional responses and more controlled cognitive responses play crucial and, in some cases, mutually competitive roles. More specifically, this theory associates controlled cognitive processing with utilitarian (or consequentialist) moral judgment aimed at promoting the "greater good" (Mill, 1861/1998). In contrast, this theory associates intuitive emotional processing with deontological judgment aimed at respecting rights, duties, and obligations (Kant, 1785/1959) that may trump the greater good.

We developed this theory in response to a longstanding philosophical puzzle known as the trolley problem (Foot, 1978; Thomson, 1985; Fischer & Ravizza, 1992). First, consider the following moral dilemma, which we will here call the *switch* case (Thomson, 1985): A runaway trolley is about to run over and kill five people, but you can save them by hitting a switch that will divert the trolley onto a side track, where it will run over and kill only one person. Here, most people say that it is morally acceptable to divert the trolley (Petrinovich, O'Neill, & Jorgensen, 1993), a judgment that accords well with the utilitarian perspective emphasizing the greater good. In the contrasting *footbridge* dilemma, a runaway trolley once again threatens five people. Here, the only way to save the five is to push a large person off a footbridge and into the trolley's path, stopping the trolley but killing the person who is pushed. (You are too small to stop the trolley yourself.) Here, most people say that it is wrong to trade one life for five, consistent with the deontological perspective, according to which individual rights often trump utilitarian considerations.

We hypothesized that people tend to disapprove of the action in the *footbridge* dilemma because the harmful action in that case, unlike the action in the *switch* case, elicits a prepotent negative emotional response that inclines people toward disapproval (figure 68.2*E*). We hypothesized further that people tend to approve of the action in the *switch* case because, in the absence of a countervailing prepotent emotional response, they default to a utilitarian mode of reasoning that favors trading one life for five (figure 68.2*A*). We proposed that the negative emotional response elicited by the *footbridge* case is related to the more "personal" nature of the harm in that case. We proposed, in other words, there is an emotional appraisal process (Scherer, Schorr, & Johnstone, 2001) that distinguishes personal dilemmas such as the *footbridge* case from impersonal dilemmas such as the *switch* case (figure 68.2*D*).

To test these hypotheses we devised a set of "personal" dilemmas modeled loosely on (and including) the *footbridge* case and a contrasting set of "impersonal" dilemmas modeled loosely on (and including) the *switch* case.[1] The effects of these stimuli were compared by using fMRI. As was predicted, the personal dilemmas preferentially engaged brain regions associated with emotion, including the mPFC, PCC, and the amygdala (Greene et al., 2001, 2004). (As was noted above, this contrast also revealed preferential engagement of the pSTS/TPJ.) Also consistent with our dual-process theory, the impersonal moral dilemmas, relative to "personal" ones, elicited increased activity in regions of DLFPC associated with working memory (Cohen et al., 1997; Smith & Jonides, 1997) and cognitive control (Miller & Cohen, 2001).

According to the dual-process theory, the *footbridge* dilemma elicits a conflict between utilitarian reasoning and emotional intuition in which the latter tends to dominate. In other cases, these opposing forces appear to be more balanced. Consider the *crying baby* dilemma: It is wartime. You and your fellow villagers are hiding from nearby enemy soldiers in a basement. Your baby starts to cry, and you cover your baby's mouth to block the sound. If you remove your hand, your baby will cry loudly, and the soldiers will hear. They will find you, your baby, and the others, and they will kill all of you. If you do not remove

your hand, your baby will smother to death. Is it morally acceptable to smother your baby to death to save yourself and the other villagers?

Here, people are relatively slow to respond and exhibit no consensus in their judgments (Greene et al., 2004). According to the dual-process theory, these behavioral effects are the result of the aforementioned conflict between emotional intuition and controlled cognition. This theory makes two key predictions. First, if dilemmas such as the *crying baby* dilemma elicit response conflict, then we would expect these dilemmas (as compared to personal dilemmas that elicit shorter RTs and less disagreement) to be associated with increased activity in the ACC, a region that is known for its sensitivity to response conflict (Botvinick, Braver, Barch, Carter, & Cohen, et al., 2001) (see Figure 68.2*C*). Second, if making utilitarian judgments in such cases requires overriding a prepotent, countervailing emotional response, then we would expect such judgments to be associated with increased activity in regions of DLPFC associated with cognitive control (Greene et al., 2001, 2004) (see Figure 68.2*B*). Both of these predictions were confirmed (Greene et al., 2004).

Three more recent studies support the dual-process theory by indicating a causal relationship between emotional responses and deontological/nonutilitarian moral judgments. Mendez, Anderson, and Shapira (2005) found that patients with frontotemporal dementia, who are known for their "emotional blunting," were disproportionately likely to approve of the action in the *footbridge* dilemma. Koenigs and colleagues (Koenigs et al., 2007) and Ciaramelli, Muccioli, Ladavas, and di Pellegrino (2007) observed similar results in patients with emotional deficits due to VMPFC lesions. The results of the former study, which distinguished high-conflict personal dilemmas such as the *crying baby* dilemma from low-conflict personal dilemmas, were particularly dramatic (see figure 68.3). In each of the high-conflict dilemmas, the VMPFC patients gave more utilitarian judgments than the control subjects did. Finally, Valdesolo and DeSteno (2006) found that normal participants were more likely to approve of the action in the *footbridge* dilemma following a positive emotion induction aimed at counteracting negative emotional responses.

Four other studies support the link between utilitarian judgment and controlled cognition. My colleagues and I conducted a cognitive load study in which subjects responded to high-conflict personal dilemmas while performing a secondary task (detecting presentations of the number "5" in a stream of numbers) designed to interfere with controlled cognitive processes. The cognitive load manipulation slowed down utilitarian judgments but had no effect on RT for deontological/nonutilitarian judgments, consistent with the hypothesis that utilitarian judgments, unlike deontological judgments, are preferentially supported by controlled cognitive processes (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008).

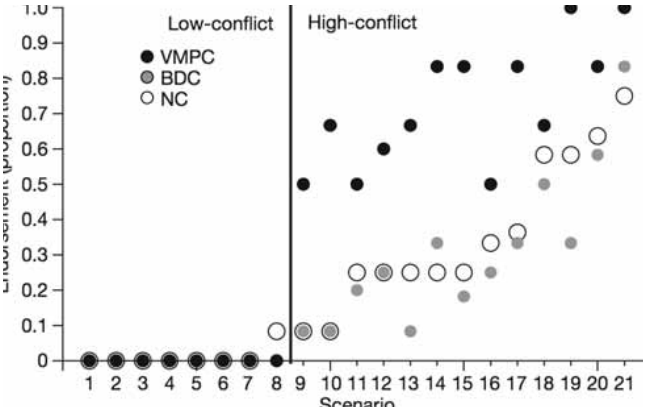


FIGURE 68.3 Frequency of utilitarian judgment in response to 21 personal moral dilemmas for VMPFC patients (black), brain-damaged controls (gray), and normal controls (white). Individual dilemmas are ordered from left to right along the *x*-axis in order of increasing frequency of utilitarian judgment among the normal controls. Dilemmas to the left of the vertical line are low-conflict and did not elicit group differences in moral judgment. Dilemmas to the right are high-conflict. The VMPFC group made a higher proportion of utilitarian judgments than both control groups in response to each of the high-conflict personal moral dilemmas. (Figure from Koenigs et al., 2007, courtesy of Nature Publish Group.)

The three remaining studies examined the relationship between moral judgment and individual differences in cognitive style. Bartels (2008) found that individuals who are high in "need for cognition" (Cacioppo, Petty, & Cao, 1984) and low in "faith in intuition" (Epstein, Pacini, DenesRaj, & Heier, 1996) were more utilitarian. Along similar lines, Hardman (2008) examined moral judgment using the Cognitive Reflection Test (Frederick, 2005), which asks people questions like this: "A bat and a ball cost $1.10. The bat costs one dollar more than the ball. How much does the ball cost?" The intuitive answer is 10¢, but a moment's reflection reveals that the correct answer is 5¢. The people who correctly answered these questions were about twice as likely to give utilitarian responses to the *footbridge* and *crying baby* dilemmas. Finally, Moore, Clarke, and Kane (2008) found that individuals with higher working memory capacity were more likely to give utilitarian judgments in response to dilemmas in which harm to the victim is inevitable. Note, however, that Killgore and colleagues (2007) found that sleep deprivation made subjects more utilitarian. This effect, however, was not observed in individuals who were high in emotional intelligence, suggesting the operation of complex emotion-cognition interactions that are not readily explained by current theory.

Three more recent fMRI studies support and broaden the dual-process theory. My colleagues and I compared dilemmas such as the *switch* case to similar dilemmas in

which saving the five people requires breaking a promise (e.g., the agent had promised the potential victim that he would not be run over). In these cases, it is the harm's *social structure*, rather than its *physical structure*, that generates the tension between utilitarian and deontological judgment. We found, first, that introducing the promise factor reproduces the familiar pattern of mPFC/PCC/TPJ activity and, second, that utilitarian judgment in the promise dilemmas is associated with increased activity in the DLPFC (right BA 46) (Greene, et al., 2009c). In a second study (Greene, et al., 2009b), we compared utilitarian and nonutilitarian/deontological moral disapproval. The *footbridge* dilemma typically elicits deontological disapproval ("It is wrong to kill the one, despite the greater good"). One can generate utilitarian disapproval using dilemmas such as the *reverse switch* case, in which one can divert the trolley onto five people in order to save one person (an action that makes no utilitarian sense). Consistent with the dual-process theory, we found that utilitarian disapproval, as compared to deontological disapproval, was associated with greater activity in the same region of DLPFC, as above. It is worth noting that the region of DLPFC that is associated with utilitarian judgment in these studies (BA 46) is posterior to that associated with utilitarian judgment in response to high-conflict personal moral dilemmas (BA 10) (Greene et al., 2004). All utilitarian judgments appear to require utilitarian reasoning, but additional cognitive control is only required in the face of countervailing emotional responses. Thus it is possible that BA 46 is engaged during utilitarian moral reasoning, while BA 10 is engaged in the more extended cognitive processing elicited by high-conflict personal dilemmas. Finally, as was noted above, Glenn and colleagues (2008) found that individuals with high psychopathy scores exhibited reduced amygdala activity during the contemplation of personal moral dilemmas, thus providing further evidence for the connection between emotion and deontological impulses (which are reliably generated by personal moral dilemmas). They also found that individuals with high scores on the interpersonal factor of the Psychopathy Checklist (which involves manipulation, conning, superficiality, and deceitfulness) (Hare, 2003) exhibited decreased activation in the mPFC/PCC/TPJ network. (Note, however, that the psychopaths did not exhibit abnormal moral judgment behavior, complicating this interpretation.)

In sum, the dual-process theory of moral judgment, which emphasizes both emotional intuition and controlled cognition, is supported by multiple fMRI studies using different behavioral paradigms, multiple behavioral studies of neurological patients, and a variety of behavioral studies using both experimental manipulations and individual difference measures. (For an alternative perspective, see Moll & de Oliveira-Souza (2007). For my reply, see Greene (2007a).)

The mental states of moral agents

As Oliver Wendell Holmes, Jr. famously observed, even a dog knows the difference between being tripped over and being kicked. Holmes's comment highlights the importance of information concerning the mental states of moral agents and, more specifically, the distinction between intentional and accidental harm.

Berthoz and colleagues (2006) identified several brain regions that exhibit increased activity in response to intentional (versus accidental) moral transgressions, including the amygdala, the precuneus, the ACC, and the DLPFC. These results suggest a kind of dual-process response to intentional harms. That may be correct, but a more recent set of studies complicates this picture. Young and colleagues (2007) compared neural responses to intended harms, accidental harms, failed attempted harms, and ordinary harmless actions (a 2×2 design crossing mental state information (agent did/did not anticipate harm) and outcome information (harm did/did not result)). They found that that the mPFC, PCC, and TPJ, all regions associated with theory of mind (Saxe, Carey, & Kanwisher, 2004), were sensitive not only to belief (i.e. anticipation) information, but also to the interaction between belief and outcome information. More specifically, the right TPJ was particularly sensitive to attempted harm, consistent with the behavioral finding that attempted harm is readily condemned, while accidental harm is not so readily excused (see also Cushman, Young, & Hauser, 2006). Interestingly, Young and colleagues (2007) found that judgments in response to accidental harm (as compared to intentional harm) were associated with increased activity in the ACC and DLPFC, regions that are associated respectively with conflict and control in the context of moral judgment (Greene et al., 2004). Young and colleagues argue that this is due to a conflict between an outcome-based response (the person caused harm) and one based on mental states (it was an accident). Thus we see here increased activity suggestive of cognitive conflict and control in response to accidental harms as opposed to intentional harms (Berthoz et al., 2006).

Further studies by Young and Saxe have examined the roles of various neural regions in processing mental state information in the context of moral judgment. They have found that the mPFC is sensitive to the valence of the agent's belief, while the TPJ and precuneus appear to be critical for the encoding and subsequent integration of belief information in moral judgment (Young & Saxe, 2008). A third study (Young & Saxe, 2009) suggests that the right TPJ, PCC, and mPFC are involved in the generation of spontaneous mental state attributions. Finally, they found, as predicted, that disrupting activity in the right TPJ using TMS produces a more childlike pattern of moral judgment (Piaget, 1965) based more on outcomes and less on mental state

information (Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2008).

While most humans (and perhaps some canines) are explicitly aware of the distinction between intended and accidental harm, people's judgments are also sensitive to a more subtle distinction between harms that are intended as a means to an end and harms that are merely foreseen as side effects (Aquinas, 2006). The means/side effect distinction is, in fact, a key distinction that distinguishes the *footbridge* dilemma (in which a person is used as a trolley-stopper) from the *switch* dilemma (in which the a person is killed as "collateral damage") (Foot, 1978; Thomson, 1985; Mikhail, 2000; Cushman et al., 2006; Moore et al., 2008). (Recent research suggests that the means/side effect distinction interacts with factors related to "personalness" in generating the effects that give rise to the trolley problem (Greene, et al., 2009a). Schaich Borg, Hynes, van Horn, Grafton, Hynes, and Sinnott-Armstrong (2006) found that the anterior STS and VMPFC exhibit increased activity in response to dilemmas in which the harm is an intended means, as opposed to a foreseen side effect. They also found increased DLPFC activity associated with harms caused through action, as opposed to inaction, consistent with the finding that people appear to have conscious access to the action/inaction distinction (Cushman et al., 2006).

While the studies described above highlight the importance of mental state representation in moral judgment, a study of moral judgment in autistic children indicates that some basic moral judgments do not depend on theory of mind abilities (Leslie, Mallon, & DiCorcia, 2006).

Neuroeconomics

Morality, broadly construed, may be viewed as a set of psychological adaptations that allow individuals to reap the benefits of cooperation (Darwin, 1871/2004). In economics, the most widely used experimental paradigm for studying cooperation is the prisoner's dilemma (Axelrod & Hamilton, 1981), in which two individuals maximize their total payoffs by cooperating but maximize their individual payoffs by not cooperating ("defecting"). Rilling and colleagues (2007) found that brain regions associated with reward (nucleus accumbens, caudate nucleus, VMPFC and OFC, and rostral ACC) were associated with cooperation, indicating that cooperative behavior is supported by general-purpose reward circuitry. A more recent study (Moll et al., 2006) in which people made charitable donations from inside the scanner teaches a similar lesson. Decisions to make costly donations were associated with increased activity in reward-related brain regions overlapping with those identified by Rilling and colleagues (2007). This study also found a remarkably high correlation ($r = .87$) between self-reported engagement in voluntary activities and the level of activation in the mPFC during costly donation.

Several neuroeconomic experiments have used the Ultimatum Game (UG) (Guth, Schmittberger, & Schwarze, 1982) to examine neural responses to unfairness. In the UG, one player (the proposer) makes a proposal about how to divide a fixed sum of money between herself and the other player (the responder). The responder may either accept the proposal, in which case the money is divided as proposed, or reject it, leaving both players with nothing. Responders typically reject offers that are substantially below half of the total as unfair. Sanfey, Rilling, Aronson, Nystrom, and Cohen (2003) found that responders responded to such unfair offers with increased activity in the insula, which is associated with autonomic arousal (Critchley, Elliott, Mathias, & Dolan, 2000) and negative emotion (Calder et al., 2001). The level of insula activity scaled with the magnitude of the unfairness, responded more to human versus computer-generated proposals and was associated with higher levels of rejection. Unfair offers also elicited increased activity in the right DLPFC, which was interpreted as being involved in inhibiting the negative emotional response to unfairness. A more recent study (Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006), however, challenges this interpretation, finding that disrupting activity in the right DLPFC generated fewer rejections of unfair offers. These results suggest that the right DLPFC is involved in inhibiting the appetitive desire for more money rather than the punitive response to unfair treatment. Koenigs and colleagues (Koenigs & Tranel, 2007) found that patients with VMPFC damage exhibited the opposite behavioral pattern, suggesting that the VMPFC plays a critical role in regulating the emotional response that drives individuals to respond punitively to unfair treatment. (Increased rejection rates can also be generated by decreasing seratonin levels through tryptophan-depletion (Crockett, Clark, Tabibnia, Lieberman, & Robbins, 2008).) A more recent fMRI study of the UG (Tabibnia, Satpute, & Lieberman, 2008) found that increased activity in the ventrolateral PFC is correlated with increased acceptance of unfair offers, suggesting that this region may play the role originally attributed to the right DLPFC.

Other neuroeconomic studies have focused on how individuals track and respond to the moral status of others. Singer, Kiebel, Winston, Dolan, and Frith (2004) examined neural responses to faces of people who had played either fairly (i.e., cooperatively) or unfairly in a sequential prisoner's dilemma game. Surprisingly, they found that faces of fair players, but not unfair players, elicited increased activity in the insula and the amygdala, regions that are widely, but not exclusively, associated with negative affect (Adolphs, 1999). In a second study, Singer and colleagues (2006) exam-

ined the interaction between (un)fair behavior and empathy. Both males and females exhibited signs of pain-empathy (increased activity in the frontoinsular cortex and ACC) when observing fair players receive a painful shock, but this effect was significantly reduced in males when the players receiving the shock had played unfairly. Males, moreover, exhibited increased reward-related activity in the nucleus accumbens (correlated with self-reported desire for revenge) when observing unfair players getting shocked. In a similar vein, de Quervain and colleagues (2004) observed that reward-related activity in the caudate nucleus was associated with willingness to punish individuals who betrayed the subject's trust in a trust game. (A trust game is essentially a sequential prisoner's dilemma game in which cooperators must trust one another to continue cooperation.) A study by Delgado, Frank, and Phelps (2005) examined the effect of reputation on moral-economic interaction. They had subjects play a trust game with fictional individuals who were characterized as good, bad, or neutral on the basis of their personal histories. The individuals' reputations affected subjects' willingness to trust them and modulated the level of activity in the caudate nucleus, partially overriding the effect of feedback during the game. King-Casas and colleagues (2005) used a trust game to examine the temporal dynamics of trust development. They found that reward-related signals in the dorsal striatum were associated with the intention to trust and were shifted earlier in time as trust developed over the course of the game, mirroring effects that are observed in nonsocial reinforcement learning (Schultz, Dayan, & Montague, 1997). Taking a more molecular approach to the understanding of trust, Kosfeld, Heinrichs, Zak, Fischbacher, and Fehr (2005) found that intranasal administration of oxytocin, a neuropeptide that is known for its role in social attachment and affiliation in nonhuman mammals (Insel & Young, 2001), increased trusting behavior.

Hsu, Anen, and Quartz (2008) examined the neural bases of decisions concerning distributive justice, pitting deontological considerations for equality against utilitarian considerations in favor of maximizing aggregate benefits ("efficiency"). Their subjects allocated money to children in an orphanage, with some options favoring equality at the expense of efficiency and vice versa. Aversion to inequality was associated with increased activity in the insula, while activity in the putamen was positively correlated with the efficiency of the outcome. The caudate nucleus, in contrast, was sensitive to both factors, reflecting the subjective utility of the option. While at odds with the relatively simple dual-process theory presented above, these results are consistent with Hume's (1739/1978) conjecture (Greene et al., 2004; Greene, 2007b) that both deontological and utilitarian considerations ultimately have affective bases, despite the latter's greater dependence on controlled cognitive processing.

Conclusion

People often speak of a "moral sense" or a "moral faculty" (Hauser, 2006), but there is no single system within the human brain that answers to this description. Rather, moral judgment emerges from a complex interaction among multiple neural systems whose functions are typically not (and might not ever be) specific to moral judgment (Greene & Haidt, 2002). The bulk of the research discussed above rightly emphasizes the role of emotion, in all of its functional and anatomical variety. At the same time, it is clear that controlled cognitive processing plays an important role in moral judgment, particularly in supporting judgments that run counter to prepotent emotional responses.

Three positive trends emerge from the foregoing discussion: First, we have seen a shift away from purely stimulus-based studies in favor of studies that associate patterns of neural activity with behavior. Second, and relatedly, we have seen an increased reliance on behavioral data, both in neuroscientific research and in complementary behavioral studies. Third, we have developed more ecologically valid paradigms involving real decisions, while recognizing that more stylized, hypothetical decisions can, like the geneticist's fruit fly, teach us valuable lessons. With regard to this issue, it is worth noting that in modern democracies, our most important decisions are made indirectly by voters whose individual choices have little bearing on outcomes and are thus effectively hypothetical.

Our current neuroscientific understanding of moral judgment is rather crude, conceptualized at the level of gross anatomical brain regions and psychological processes that are familiar from introspection. But for all our ignorance, the physical basis of moral judgment is no longer a complete mystery. We not only have identified brain regions that are "involved" in moral judgment, but also have begun to carve the moral brain at its functional joints.

ACKNOWLEDGMENTS Many thanks to Joe Paxton and Liane Young for their helpful comments.

NOTE

1. We defined "personal" moral dilemmas/harms as those involving actions that are (1) likely to cause serious bodily harm (2) to a particular person and (3) this harm does not result from deflecting an existing threat onto a different party (Greene et al., 2001). The first two criteria exclude minor harms and harms to indeterminate "statistical" individuals, respectively. The third criterion aims to capture a sense of agency, distinguishing between harms that are "authored" rather than merely "edited" by the agent in question. Recent research suggests that the dilemmas that were originally classified as "personal" and "impersonal" can be fruitfully classified in other ways (Mikhail, 2000; Royzman & Baron, 2002; Cushman et al., 2006; Waldmann & Dieterich, 2007; Moore et al., 2008; Greene et al., 2009a).

REFERENCES

Adolphs, R. (1999). Social cognition and the human brain. *Trends Cogn. Sci.*, *3*, 469–479.

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.*, *4*, 165–178.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nat. Rev. Neurosci.*, *7*, 268–277.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat. Neurosci.*, *2*, 1032–1037.

Aquinas, T. (2006). *Summa theologiae.* Cambridge, UK: Cambridge University Press.

Axelrod, R., & Hamilton, W. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.

Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*, 381–417.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cereb. Cortex*, *6*, 215–225.

Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D., & Knight, R. T. (2003). The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *J. Pers. Soc. Psychol.*, *85*, 594–604.

Berkowitz, L. (1993). *Aggression: Its causes, consequences and control.* Philadelphia: Temple University Press.

Berthoz, S., Armony, J. L., Blair, R. J., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain*, *125*, 1696–1708.

Berthoz, S., Grezes, J., Armony, J. L., Passingham, R. E., & Dolan, R. J. (2006). Affective response to one's own moral violations. *NeuroImage*, *31*, 945–950.

Blair, R. J. (1995). A cognitive developmental approach to mortality: Investigating the psychopath. *Cognition*, *57*, 1–29.

Blair, R. J. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *J. Neurol. Neurosurg. Psychiatry 71*, 727–731.

Blair, R. J. (2004). The roles of orbital frontal cortex in the modulation of antisocial behavior. *Brain Cogn.*, *55*, 198–208.

Blair, R. J. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends Cogn. Sci.*, *11*, 387–392.

Blair, R. J, Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, *34*, 192–198.

Blonigen, D. M., Hicks, B. M., Krueger, R. F., Patrick, C. J., & Iacono, W. G. (2005). Psychopathic personality traits: Heritability and genetic overlap with internalizing and externalizing psychopathology. *Psychol. Med.*, *35*, 637–648.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol. Rev.*, *108*, 624–652.

Buckner, R. L., & Carrol, D. C. (2007). Self-projection and the brain. *Trends Cogn. Sci.*, *11*, 49–57.

Cacioppo, J., Petty, R., & Kao, C. (1984). The efficient assessment of need for cognition. *J. Pers. Assessment*, *48*, 306–307.

Calder, A. J., Lawrence, A. D., & Young, A. W. (2001). Neuropsychology of fear and loathing. *Nat. Rev. Neurosci. 2*, 352–363.

Chaiken, S., & Trope, Y., (Eds.). (1999). *Dual-process theories in social psychology*. New York: Guilford.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Soc. Cogn. Affect. Neurosci.*, *2*, 84–92.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., et al. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*, 604–608.

Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment.* New York: Cambridge University Press.

Critchley, H. D., Elliott, R., Mathias, C. J., & Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study. *J. Neurosci.*, *20*, 3033–3040.

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., & Robbins, T. W. (2008). Serotonin modulates behavioral reactions to unfairness. *Science*, *320*, 1739.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychol. Sci. 17*, 1082–1089.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: G.P. Putnam.

Darwin, C. (1871/2004). *The descent of man.* New York: Penguin.

Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Mol. Psychiatry 6*, 13–34.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–1258.

de Waal, F. (2006). *Primates and philosophers: How morality evolved*. Princeton, NJ: Princeton University Press.

Delgado, M. R., Frank, R., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.*, *8*, 1611–1618.

Diergaarde, L., Gerrits, M. A., Brouwers, J. P., & van Ree, J. M. (2005). Early amygdala damage disrupts performance on medial prefrontal cortex-related tasks but spares spatial learning and memory in the rat. *Neuroscience*, *130*, 581–590.

Epstein, S., Pacini, R., DenesRaj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *J. Pers. Soc. Psychol.*, *71*, 390–405.

Finger, E. C., Marsh, A. A., Kamel, N., Mitchell, D. G., & Blair, J. R. (2006). Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behavior. *NeuroImage*, *33*, 414–421.

Fischer, J. M., & Ravizza, M., (Eds.). (1992). *Ethics: Problems and principles*. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.

Foot, P. (1978). The problem of abortion and the doctrine of double effect. In *Virtues and vices* (pp. 19–32). Oxford, UK: Blackwell.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically orgainzed into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. USA*, *102*, 9673–9678.

Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspectives*, *19*, 25–42.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron*, *32*, 969–979.

Glenn, A., Raine, A., & Schug, R. (2008). The neural correlates of moral decision-making in psychopathy. *Mol. Psychiatry*, *14*, 5–6.

Grattan, L. M., & Eslinger, P. J. (1992). Long-term psychological consequences of childhood frontal lobe lesion in patient DT. *Brain Cogn. 20*, 185–195.

Greene, J. D. (2007a). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends Cogn. Sci.*, *11*, 322–323; author reply 323–324.

Greene, J. D. (2007b). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology: Vol. 3. The neuroscience of morality: Emotion, disease, and development.* Cambridge, MA: MIT Press.

Greene, J. D. (in press). Social neuroscience and the soul's last stand. In A. Todorov, S. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind.* New York: Oxford University Press.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom. L. E., & Cohen, J. D. (2009a). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends Cogn. Sci.*, *6*, 517–523.

Greene, J. D., Lowenberg, K., Paxton, J., Nystrom, L. E., & Cohen, J. D. (2009b). Neural dissociation between affective and cognitive moral disapproval.

Greene, J., D., Lowenberg, K., Paxton, J., Nystrom, L., Darley, J., & Cohen, J. D. (2009c). Duty vs. the greater good: Dissociable neural bases of deontological and utilitarian moral judgment in the context of keeping and breaking promises.

Greene, J. D., Morelli, S., Lowenberg, K., Nystrom, L., & Cohen, J. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144–1154.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389–400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108.

Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proc. Natl. Acad. Sci. USA*, *98*, 4259–4264.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nat. Rev. Neurosci.*, *2*, 685–694.

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organization*, *3*, 367–388.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol. Rev.*, *108*, 814–834.

Hamlin, J., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557–560.

Hardman, D. (2008). Moral dilemmas: Who makes utilitarian choices. Unpublished manuscript.

Hare, R. D. (1991). *The Hare psychopathy checklist—Revised.* Toronto: Multi-Health Systems.

Hare, R. (2003). *Hare psychopathy checklist—Revised (PCL-R)* (2nd ed.). Toronto: Multi-Health Systems.

Harenski, C. L., & Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*, *30*, 313–324.

Hauser, M. (2006). The liver and the moral organ. *Soc. Cogn. Affect. Neurosci.*, *1*, 214–220.

Heekeren, H. R., Wartenburger, I., Schmidt, H., Prehn, K., Schwintowski, H. P., & Villringer, A. (2005). Influence of bodily harm on neural correlates of semantic and moral decision-making. *NeuroImage*, *24*, 887–897.

Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *NeuroReport*, *14*, 1215–1219.

Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, *320*, 1092–1095.

Hume, D. (1739/1978). *A treatise of human nature* (2nd ed.), edited by L. A. Selby-Bigge & P. H., Nidditch (pp. xix, 743). Oxford, UK: Oxford University Press.

Insel, T. R., & Young, L. J. (2001). The neurobiology of attachment. *Nat. Rev. Neurosci.*, *2*, 129–136.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *Am. Psychol.*, *58*, 697–720.

Kant, I. (1785/1959). *Foundation of the metaphysics of morals.* Indianapolis: Bobbs-Merrill.

Kedia, G., Berthoz, S., Wessa, M., Hilton, D., & Martinot, J. (2008). An agent harms a victim: A functional magnetic imaging study of specific moral emotions. *J. Cogn. Neurosci.*, *20*, 1788–1798.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati. S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *J. Cogn. Neurosci.*, *14*, 785–794.

Kiehl, K. A. (2006). A cognitive neuroscience perspective on psychopathy: Evidence for paralimbic system dysfunction. *Psychiatry Res.*, *142*, 107–128.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., et al. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biol. Psychiatry*, *50*, 677–684.

Killgore, W. D., Killgore, D. B., Day, L. M., Li, C., Kamimori, G. H., & Balkin, T. J. (2007). The effects of 53 hours of sleep deprivation on moral judgment. *Sleep*, *30*, 345–352.

King, J. A., Blair, R. J., Mitchell, D. G., Dolan, R. J., & Burgess, N. (2006). Doing the right thing: A common neural circuit for appropriate violent or compassionate behavior. *NeuroImage*, *30*, 1069–1076.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R., (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, *308*, 78–83.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*, 829–832.

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the Ultimatum Game. *J. Neurosci.*, *27*, 951–956.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*, 908–911.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A., Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago: Rand McNally.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*, 673–676.

Leslie, A., Mallon, R., & DiCorcia, J. (2006). Transgressors, victims, and cry babies: Is basic moral judgment spared in autism? *Soc. Neurosci*, *1*, 270–283.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Adv. Exp. Soc. Psychol.*, *34*, 199–249.

Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., & Blair, R. J. (2006). The neural basis of implicit moral attitude: An IAT study using event-related fMRI. *NeuroImage*, *30*, 1449–1457.

Macmillan, M. (2000). *An odd kind of fame*. Cambridge, MA: MIT Press.

Maddock, R. J. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends Neurosci.*, *22*, 310–316.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science*, *315*, 393–395.

Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cogn. Behav. Neurol.*, *18*, 193–197.

Mikhail, J. (2000). *Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in* A Theory of Justice. Unpublished Ph.D. dissertation, Cornell University.

Mill, J. S. (1861/1998). *Utilitarianism*, edited by R. Crisp. New York: Oxford University Press.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.*, *24*, 167–202.

Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends Cogn. Sci.*, *11*, 319–321.

Moll, J., de Oliveira-Souza, R., Bramati, I., & Grafman, J. (2002a). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage*, *16*, 696–703.

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., et al. (2002b). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *J. Neurosci.*, *22*, 2730–2736.

Moll, J, de Oliveira-Souza, R, Moll, F., Ignacio, F., Bramati, I., Caparelli-Daquer, E., et al. (2005). The moral affiliations of disgust: A functional MRI study. *Cogn. Behav. Neurol.*, *18*, 68–78.

Moll, J., Eslinger, P. J., & Oliveira-Souza, R. (2001). Frontopolar and anterior temporal cortex activation in a moral judgment task: Preliminary functional MRI results in normal subjects. *Arq. Neuropsiquiatr.*, *59*, 657–664.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. USA*, *103*, 15623–15628.

Moore, A., Clark, B., & Kane, M. (2008). Who shalt not kill?: Individual differences in working memory capacity, executive control, and moral judgment. *Psychol. Sci.*, *19*, 549–557.

Muller, J. L., Sommer, M., Dohnel, K., Weber, T., Schmidt-Wilcke, T., & Hajak, G. (2008). Disturbed prefrontal and temporal brain function during emotion and cognition interaction in criminal psychopathy. *Behav. Sci. Law*, *26*, 131–150.

Oxford, M., Cavell, T. A., & Hughes, J. N. (2003). Callous/unemotional traits moderate the relation between ineffective parenting and child externalizing problems: A partial replication and extension. *J. Clin. Child Adolesc. Psychol.*, *32*, 577–585.

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *J. Pers. Soc. Psychol.*, *64*, 467–478.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, *16*, 331–348.

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*, 495–498.

Piaget, J. (1965). *The moral judgement of the child*. New York: Free Press.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 55–85). Hillsdale, NJ: Lawrence Erlbaum.

Raine, A., Buchsbaum, M. S., Stanley, J., Lottenberg, S., Abel, L., & Stoddard, J. (1994). Selective reductions in prefrontal glucose metabolism in murderers. *Biol. Psychiatry*, *36*, 365–373.

Raine, A., Lencz, T., Bihrle, S., LaCasse, L., & Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Arch. Gen. Psychiatry 57*, 119–127; discussion, 128.

Raine, A., & Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Soc. Cogn. Affect. Neurosci.*, *1*, 203–213.

Rilling, J., Glenn, A., Jairam, M., Pagnoni, G., Goldsmith, D., Elfenbein, H., et al. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biol. Psychiatry*, *61*, 1260–1271.

Robertson, D., Snarey, J., Ousley, O., Harenski, K., DuBois Bowman, F., Gilkey, R., et al. (2007). The neural processing of moral sensitivity to issues of justice and care. *Neuropsychologia*, *45*, 755–766.

Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Soc. Justice Res.*, *15*, 165–184.

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *J. Pers. Soc. Psychol.*, *66*, 870–881.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *J. Pers. Soc. Psychol.*, *76*, 574–586.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, *300*, 1755–1758.

Saver, J., & Damasio, A. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia*, *29*, 1241–1249.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annu. Rev. Psychol.*, *55*, 87–124.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nat. Rev. Neurosci.*, *8*, 657–661.

Schaich, Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *J. Cogn. Neurosci.*, *18*, 803–817.

Schaich, Borg, J., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *J. Cogn. Neurosci.*, *20*, 1–19.

Scherer, K., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion*. New York: Oxford University Press.

Schneider, F., Habel, U., Kessler, C., Posse, S., Grodd, W., & Muller-Gartner, H. W. (2000). Functional imaging of conditioned aversive emotional responses in antisocial personality disorder. *Neuropsychobiology*, *42*, 192–201.

Schoenbaum, G., & Roesch, M. (2005). Orbitofrontal cortex, associative learning, and expectancies. *Neuron*, *47*, 633–636.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Shin, L., Dougherty, D., Orr, S., Pitman, R., Lasko, M., Macklin, M., et al. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biol. Psychiatry*, *48*, 43–50.

Singer, T., Kiebel, S., Winston, J., Dolan, R., & Frith, C. (2004). Brain response to the acquired moral status of faces. *Neuron, 41*, 653–662.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*, 466–469.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cogn. Psychol.*, *33*, 5–42.

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol. Sci.*, *19*, 339–347.

Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: An fMRI study. *NeuroImage*, *23*, 967–974.

Thomson, J. (1985). The trolley problem. *Yale Law J.*, *94*, 1395–1415.

Turiel, E. (2006). Thought, emotions and social interactional processes in moral development. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 7–35). Mahwah, NJ: Lawrence Erlbaum.

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychol. Sci.*, *17*, 476–477.

Viding, E., Blair, R. J., Moffitt, T. E., & Plomin, R. (2005). Evidence for substantial genetic risk for psychopathy in 7-year-olds. *J. Child Psychol. Psychiatry*, *46*, 592–597.

Vollm, B., Richardson, P., & Stirling, J. (2004). Neurobiological substrates of antisocial and borderline personality disorders: Preliminary results from an fMRI study. *Criminal Behav. Mental Health*, *14*, 39–54.

Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychol. Sci.*, *18*, 247–253.

Warneken, F., Hare, B., Melis, A. P., Hanus, D., & Tomasello, M. (2007). Spontaneous altruism by chimpanzees and young children. *PLoS Biol.*, *5*, e184.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychol. Sci.*, *16*, 780–784.

Yang, Y., Raine, A., Lencz, T., Bihrle, S., LaCasse, L., & Colletti, P. (2005). Volume reduction in prefrontal gray matter in unsuccessful criminal psychopaths. *Biol. Psychiatry*, *57*, 1103–1108.

Yang, Y., Raine, A., Narr, K., Lencz, T., & Toga, A. (2006). *Amygdala volume reduction in psychopaths*. Paper presented at the annual meeting of the Society for Research in Psychopathology.

Young, L., Camprodon, J., Hauser, M., Pascual-Leone, A., & Saxe, R. (2008). Disrupting neural mechanisms involved in belief attribution impairs moral judgment. Unpublished manuscript.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci. USA*, *104*, 8235–8240.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage*, *40*, 1912–1920.

Young, L., & Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *J. Cogn. Neurosci.*, *21*, 1396–1405.

IX

HIGHER COGNITIVE FUNCTIONS

Chapter 69 KNOWLTON AND HOLYOAK 1005

70 SUMMERFIELD AND KOECHLIN 1019

71 MARTIN 1031

72 McCLELLAND, ROGERS, PATTERSON, DILKINA, AND LAMBON RALPH 1047

73 RAICHLE 1067

74 RANGEL 1075

75 GLIMCHER 1085

76 PHELPS AND DELGADO 1093

Introduction

ELIZABETH A. PHELPS

Higher cognition encompasses a broad range of processes that allow us to take the products of more discrete cognitive functions and transform them to generate concepts, reason, and decide how to interact with the world. Given the complexity and diversity of processes that fall under the domain of higher cognition, part IX offers a wide variety of topics and approaches. One change from previous editions of *The Cognitive Neurosciences* is that the higher cognition section has been expanded to encompass both higher cognition and decision making. In previous editions, decision making was included sporadically under different section headings. With the exception of perceptual and motor decision making, the first edition did not include any contributions on decision making. In the second edition, decision making was included as a single contribution under higher cognition. The third edition was produced around the time neuroeconomics was starting to emerge as a new discipline within cognitive neuroscience. Reflecting this novelty, chapters on decision making were included in a section on perspective and new directions. Now, as interest in neuroeconomics has grown, research on the cognitive neuroscience of decision making has gained recognition as an important and exciting development in our endeavors to understand the neuroscience of human behavior. Reflecting this change, roughly half of the contributions in this section concern the processes driving decisions.

The contributions to section IX reflect three main themes and an interesting new perspective. The first two chapters focus primarily on the role of the prefrontal cortex as it represents higher cognitive functions. Knowlton and Holyoak in chapter 69 examine how the prefrontal cortex supports relational reasoning, that is, the ability to draw inferences based on relations, as opposed to specific features, a principle underlying analogies. By combining cognitive and

computational models with neuropsychological and imaging data, they isolate and describe the component processes and neural representations that underlie complex reasoning. Chapter 70 by Summerfield and Koechlin provides a global, hierarchical model of the prefrontal cortex that accounts for the integration of executive functions, including the representation of goals and selection, with motivational and evaluative information across the lateral, medial, and orbital prefrontal cortices. They outline how this model of the prefrontal cortex supports decision processes.

The next two chapters are concerned with the foundations of conceptual and semantic knowledge. How we group and organize the information we encounter to allow us to generalize across instances, avoid redundant representations, and extract general principles is fundamental to navigating a complex world. In chapter 71, Martin explores the foundations of conceptual knowledge by examining the neural systems mediating the formation, representation, and utilization of concepts about objects. In chapter 72, McClelland, Rogers, Patterson, Dilkina, and Lambon Ralph expand on the topic by introducing a computational framework for capturing the nature of semantic cognition. They explore how this model reflects the principles underlying both the development of conceptual knowledge and its disintegration with semantic dementia.

In chapter 73, Raichle provides a new perspective on how to view brain function. He argues that most functional imaging studies have a reflexive view of the brain, in which its primary purpose is to respond to external events. In contrast, he suggests that investigating the intrinsic operations of the brain may be equally important, if not more, in understanding brain function. Although this chapter sits apart from the other chapters in this section, in both the topic and the emphasis on methodological issues in functional imaging, it is hard to imagine what else intrinsic functions might entail, if not, at least in part, some components of higher cognition. In this way, the chapter by Raichle is not only an important perspective in how we approach our understanding of the human brain, but also, potentially, a contribution to understanding the cognitive neuroscience of higher cognitive processes.

Decision making, specifically from a neuroeconomic perspective, is the focus of the last three chapters. In chapter 74, Rangel uses a neuroeconomic approach to explore how we make a simple goal-directed choice between two options. By walking through the stages of a simple choice, he examines how computational approaches and human neuroscience studies can be combined to capture the complex representation of a relatively simple decision. In chapter 75, Glimcher highlights how the neuroeconomic approach has evolved by combining neuroscience research in human and nonhuman primates with economic and psychological models of decision making. He suggests that initial findings suggest a two-step neurobiological process that includes valuation and choice. The final chapter on this topic, chapter 76 by Delgado and myself, highlights the importance of emotion in decision making and reviews issues and progress in integrating affective neuroscience with neuroeconomics.

69 Prefrontal Substrate of Human Relational Reasoning

BARBARA J. KNOWLTON AND KEITH J. HOLYOAK

ABSTRACT Relational reasoning, including the distinctively human capacity to see analogies between disparate situations, requires the ability to mentally represent and manipulate the relationships among concepts. Over the past decade, studies of cognitive development, aging, and neurological disease have supported the hypothesis that the prefrontal cortex plays a critical role in relational reasoning. Analysis of the component processes of relational reasoning has motivated neuroimaging studies linking these components to distinct neural substrates. Studies using diverse reasoning tasks have converged on the conclusion that frontopolar cortex responds specifically when multiple relations must be integrated to solve a problem. Other subregions in the inferior and middle frontal gyri appear to be critical for resolving interference from distracting elements of the problem and for control of working memory. Mapping the components of relational reasoning is an essential first step toward understanding how the architecture of the prefrontal cortex supports human thinking.

Humans, more than any other species, are able to cope with novel problems that arise across a wide range of domains. This capacity depends in part on *role-based relational reasoning*—the ability to draw inferences about entities based on the roles they fill in relations, where the roles are not predictable by features of the entities and the relations cannot be reduced to roleless chunks (Penn, Holyoak, & Povinelli, 2008). A canonical example of role-based relational reasoning is reasoning by analogy, which enables detection of higher-order similarities between superficially dissimilar situations (Gick & Holyoak, 1980; Gentner, 1983; Holyoak, 2005). For example, modern medicine originated with Pasteur's development of the germ theory of disease by analogy to the known role of microorganisms in fermentation of grapes, coupled with Lister's extension of Pasteur's analogy to explain infections (Thagard, 1996). Analogy requires representing and integrating relational knowledge in working memory, while coping with interference from salient distracting information. For example, Pasteur grasped that microorganisms need to *be alive* and *in contact with* grapes in order to *cause* fermentation; by integrating these relations and mapping them onto what he knew about certain silkworm diseases, he was able to infer that germs might be their cause (and hence the diseases might be prevented by killing the germs or blocking their contact with silkworms). At the same time, he had to ignore many salient but irrelevant dissimilarities between the two analogs (e.g., silkworms are animals, grapes are not).

While prefrontal cortex (PFC) has long been associated with problem solving and fluid intelligence, in the past decade there has been a surge in work specifically relating the PFC to processes supporting relational reasoning. In 1995, Robin and Holyoak advanced the thesis that the PFC is responsible for the creation and maintenance of explicit relational representations that guide thought and action. At that time, however, few or no neuropsychological or neuroimaging data were available to directly connect relational processing with the human PFC. Moreover, although several major computational models of analogical reasoning had been developed in cognitive science (e.g., Falkenhainer, Forbus, & Gentner, 1989; Holyoak & Thagard, 1989), none even attempted to incorporate constraints based on what was known about working memory, interference control, and their neural substrates.

Today the picture is very different. Behavioral studies with children (Richland, Morrison, & Holyoak, 2006), young adults (Cho, Holyoak, & Cannon, 2007), and older adults (Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004; Viskontas, Holyoak, & Knowlton, 2005) have teased apart some of the major component processes in relational reasoning. Computational models of analogy have incorporated working-memory constraints (Halford, Wilson, & Phillips, 1998; Hummel & Holyoak, 1997, 2003). It has been shown that "virtual brain damage" in a neural-network model can simulate changes in reasoning related to normal aging (Viskontas et al., 2004), as well as reasoning deficits in patients with damage to their frontal or temporal cortices (Morrison et al., 2004). In addition to neuropsychological studies of relational reasoning conducted in our lab (Krawcyzk et al., 2008; Morrison et al., 2004; Waltz et al., 1999, 2004), neuroimaging studies from multiple labs have identified subregions of PFC associated with integrating multiple relations (Bunge, Wendelken, Badre, & Wagner, 2005; Christoff et al., 2001; Green, Fugelsang, Kraemer, Shamosh,

BARBARA J. KNOWLTON and KEITH J. HOLYOAK Department of Psychology, University of California, Los Angeles, California

& Dunbar, 2006; Kroger et al., 2002; Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997; Wendelken, Nakahbenko, Donohue, Carter, & Bunge, 2008), as well as with interference control (Cho et al., 2009).

In this chapter we discuss how a clearer picture of the neural basis of relational reasoning has emerged from current conceptions of its component processes.

Component processes of relational reasoning

Relational Integration and Interference Control As Robin and Holyoak (1995) argued, relational reasoning appears to require processes closely related to those associated with the operation of the PFC. As the name implies, role-based relational reasoning requires drawing inferences about entities based on the *roles* they fill, rather than on direct similarity. A specific task that instantiates these requirements is illustrated in figure 69.1, which depicts examples of "scene-analogy" problems developed by Richland and colleagues (2006) for use with children as young as 3–4 years. For each pair of pictures, children were asked to identify the object in the bottom picture that "goes with" the object indicated by an arrow in the top picture. In some problems, such as that shown in figure 69.1*B*, the child is confronted with a conflict between two possible answers, one relational and one based on perceptual and/or semantic

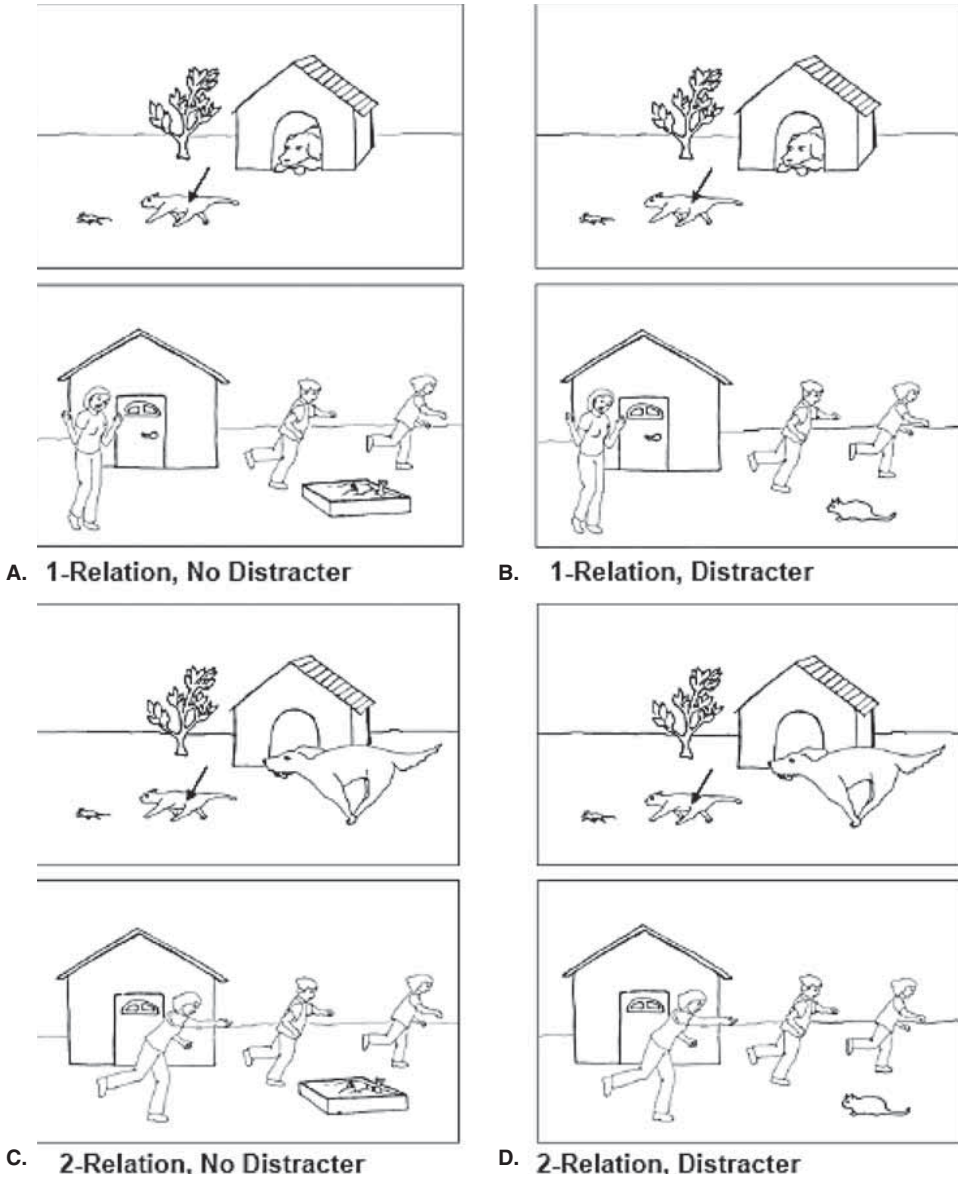


Figure 69.1 Example set of scene pairs constructed by Richland, Morrison, and Holyoak (2006). (*A*) 1-relation/no distracter; (*B*) 1-relation/distracter; (*C*) 2-relation/no distracter; (*D*) 2-relation/distracter. (Reprinted with permission from Richland, Morrison, & Holyoak, 2006.)

similarity. The cat in the top picture perceptually resembles the cat in the bottom picture, but plays a role (chasing a mouse) that parallels the role played by the boy in the bottom picture (chasing a girl). Richland and colleagues found that young children were less likely to give the relational response when an alternative based on direct similarity was available (figure 69.1*B*) than when no such distracter object was present (figure 69.1*A*, where in the bottom picture the cat has been replaced by a sandbox).

Relational reasoning varies in its complexity, which has been linked to the number of relational roles relevant to an inference (Halford, 1993; Halford et al., 1998). A critical distinction is whether a single relation is sufficient to determine the role-based inference (figure 69.1*A*,*B*), or whether multiple relations must be integrated to derive an inference (figure 69.1*C*,*D*). For example, in figure 69.1*D* it is not sufficient simply to relate the cat to another "chaser" in the bottom picture, as in the latter both the boy and the woman are chasing someone. Rather, the cat corresponds specifically to the boy because each is both being chased (by the dog and the woman, respectively) as well as chasing (the mouse and the girl, respectively). Richland and colleagues (2006) found that preschool children gave fewer relational responses when either a similar distracter was present in the bottom picture (figure 69.1*B*,*D* versus *A*,*C*) or when two relations had to be integrated (figure 69.1*C*,*D* versus *A*,*B*). By age 13–14 years—roughly the age at which the PFC has undergone substantial further maturation (Giedd, 2004)—children reliably gave the relational response even when multiple relations had to be integrated and a similar distracter was present.

LISA: A Neurocomputational Model of Relational Reasoning Research on the development of analogical reasoning thus suggests that relational reasoning depends on two major component processes: integration of multiple relations and the capacity to cope with interference from salient but relationally irrelevant information. Most computational analogies of analogy have not considered how analogical reasoning might be implemented in the brain, and traditional neural-network models of cognition encounter severe difficulty in representing explicit relations and the binding of objects into relational roles (Doumas & Hummel, 2005). However, a neural-network model of the component processes underlying relational reasoning has been developed in recent years. LISA (*Learning and Inference with Schemas and Analogies*; Hummel & Holyoak, 1997, 2003) uses a representation of knowledge based on a hierarchy of distributed and localist units (see figure 69.2). LISA's representational scheme codes relations as sets of roles (e.g., for "cat chases mouse," the cat fills the role of "chaser" and the mouse the role of "chased"). The model uses synchrony of firing to bind distributed representations of relational roles (e.g., chaser) to distributed representations of their fillers (e.g., cat). The process of "thinking about" a proposition entails keeping separate role-filler bindings firing *out* of synchrony with one another. In LISA, working memory is necessarily capacity-limited: It is only possible to keep a finite number of role-filler bindings simultaneously active and out of synchrony with one another (for details see appendix A in Hummel & Holyoak, 2003).

LISA represents propositions using a hierarchy of distributed and localist units. Figure 69.2*B* provides a schematic representation of LISA's architecture as applied to one of the scene-analogy problems used by Richland and colleagues (2006). At the bottom of the hierarchy, *semantic* units (small circles in figure 69.2*B*) represent objects and relational roles in a distributed fashion. For example, consider the

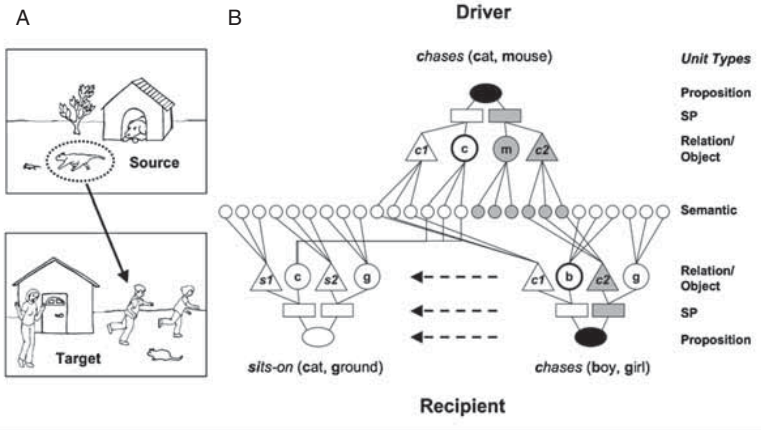


Figure 69.2 (*A*) Example of 1-relation/distracter scene-analogy problem (Richland et al., 2006). (*B*) LISA architecture as applied to this problem. In order for a reasoner to select the boy in the target as the correct analogical mapping to the cat in the source, units in the recipient representing *chases* (boy, girl) must inhibit corresponding units in the propositional structure containing the featurally similar "sitting cat" distracter. (Reprinted with permission from Morrison, Doumas, & Richland, 2006.)

proposition *chase* (cat, mouse). Each role of the *chase* relation would be represented by units coding for its semantic content (e.g., *aggressor* for the first role, *victim* for the second, and *pursuit* for both). Similarly, the objects "cat" and "mouse" would be represented by units specifying their meaning (e.g., cat: *animal, pet, soft*). *Predicate* and *object* units (triangles and large circles, respectively, in figure 69.2*B*) represent relational roles and their object fillers, and have bidirectional excitatory connections to the corresponding semantic units. Subproposition (SP) units (rectangles in figure 69.2*B*) bind roles to their arguments, and have bidirectional connections to the corresponding predicate and object units. In the case of *chase* (cat, mouse), one SP would bind "cat" to the first role of *chase*, and another would bind "mouse" to the second. At the top of the hierarchy, proposition (P) units bind role-filler bindings into complete propositions by way of excitatory connections to the corresponding SPs. A complete analog (i.e., situation, story, or scene) is represented by the collection of semantic, predicate, object, SP, and P units that collectively code the propositions in that analog. The semantic units permit the units in one analog to communicate with the units in others.

To generate an analogical mapping, units representing one analog (the *driver*) are activated in working memory, and reasoning proceeds by passing activation from these units through distributed semantic units to units representing the *recipient* analog in long-term memory. As units in the recipient analog are fired, they enter working memory. LISA postulates a set of *mapping connections* between units of the same type in separate analogs. These connections grow whenever the corresponding units are active simultaneously and thereby permit LISA to rapidly learn the correspondences between structures in separate analogs.

The basic processes of LISA are closely related to the functions of the PFC. Hummel and Holyoak (1997, 2003) hypothesized that the rapid learning of mapping connections, which is critical to relational integration, is an important function of working memory as implemented in prefrontal cortex (cf. Assad, Rainer, & Miller, 1998). Inhibitory control, which is also considered an important function of prefrontal cortex (Miller & Cohen, 2001; Shimamura, 2000), plays a central role in several aspects of LISA. These include (1) LISA's ability to select items for placement into working memory, (2) its working memory capacity for role-filler bindings, (3) its ability to control the spreading of activation in the recipient (i.e., its ability to disambiguate which elements of the recipient correspond to the active units in the driver), (4) its ability to use competition among mapping connections to enforce structural constraints on the discovery of analogical mappings, particularly the constraint that mappings tend to be one-to-one, and (5) its ability to select a relation-based response despite the availability of a salient but superficial distracter.

Importantly, LISA predicts that relational integration and inhibitory control are distinct but closely linked processes. Relational integration requires the ability to rapidly learn mapping connections, but inhibitory control is essential to set up the conditions for successful learning (i.e., simultaneous activation of objects that fill parallel relational roles). As we will see, LISA's conception of the relationship between relational integration and inhibitory control is consistent with recent functional imaging data that indicate how these processes map onto subregions of PFC.

Neuropsychological evidence for the role of PFC in relational reasoning

One of the major sources of evidence concerning the dependence of relational reasoning on the integrity of the PFC comes from neuropsychological studies of patients diagnosed with frontotemporal lobar degeneration (FTLD), a dementia subtype distinct from Alzheimer's disease (Brun, 1993; Snowden, Neary, & Mann, 2007). FTLD occurs in three variants: a progressive aphasia, in which areas involved in language production are primarily affected initially; a frontal variant, which is characterized in early stages by prefrontal atrophy; and a temporal variant (also known as *semantic dementia*) in which degeneration includes left anterior temporal cortex (Hodges, Patterson, Oxbury, & Funnell, 1996). Frontal-variant FTLD provides a model for investigating what reasoning processes are dependent on PFC, while the temporal variant provides a closely matched control with the same underlying disease.

Relational Complexity and the PFC Waltz and colleagues (1999) examined performance of FTLD patients and age-matched normal control subjects on simple reasoning tasks, using closely matched variants of problems that differed specifically in whether or not success required integration of multiple relations. They hypothesized that patients with prefrontal cortical dysfunction would exhibit impaired performance when asked to integrate multiple relations, yet would perform normally when only one relation needed to be considered. The performance of frontal-variant FTLD patients was compared to that of temporal-variant patients, as well as age-matched healthy controls.

Figure 69.3 illustrates how relational complexity was manipulated for problems adapted from the Raven's Standard Progressive Matrices Test (RPM), which has long been used as a measure of reasoning ability (Raven, 1941). Nonrelational problems (level-0 complexity) involved a visual pattern, with a blank space in the bottom right-hand corner (see figure 69.3*A*). On these problems, the participant could simply pattern-match to select the correct completion. Each one-relation problem (level-1 complexity) involved a 2×2 matrix that required processing one relational change over

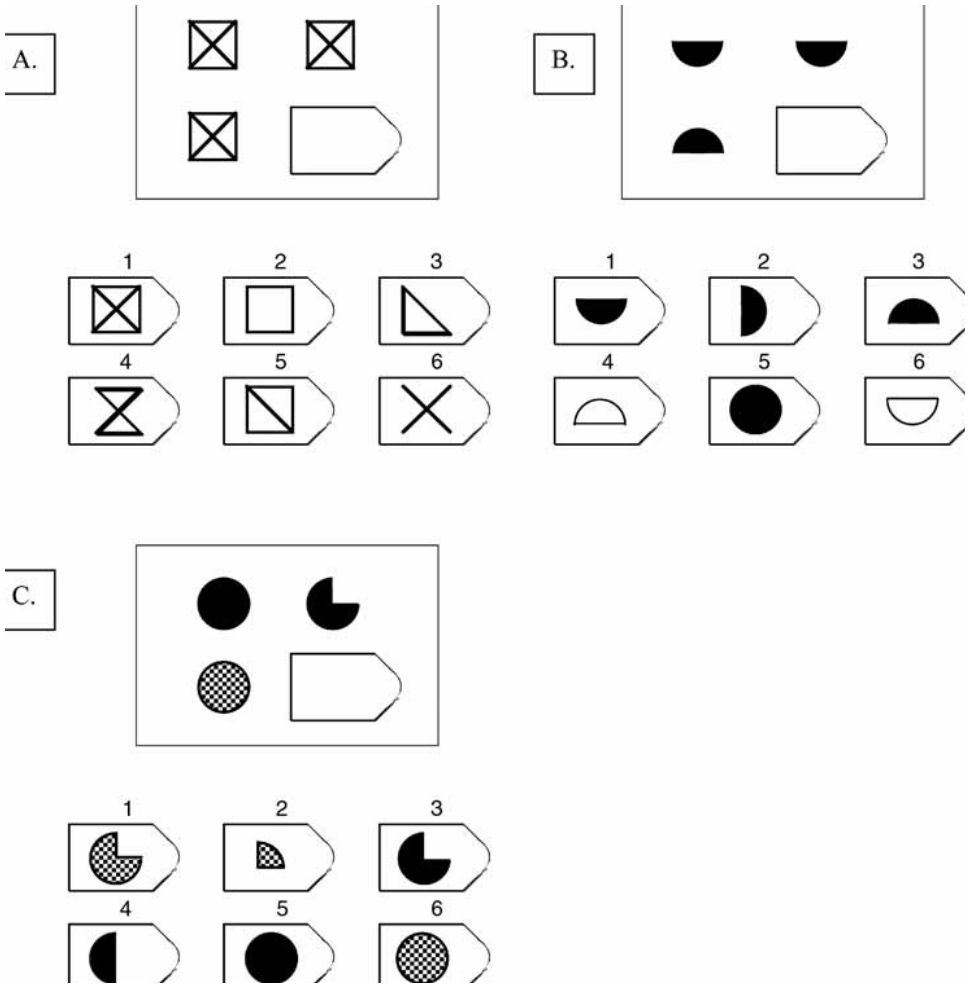


FIGURE 69.3 Examples of problems adapted from the Raven Standard Progressive Matrices Test by Waltz and colleagues (1999). (*A*) Nonrelational problem (level 0), requiring only perceptual matching (correct response is choice 1). (*B*) One-relation problem (level 1), requiring processing of the transformation along the vertical dimension only (reflection across the *x*-axis) in order to choose the correct alternative (choice 3). (*C*) Two-relation problem (level 2), requiring integration of the relation along the vertical dimension (solid to checked pattern) and the relation across the horizontal dimension (removal of the upper-right quadrant) in order to choose the correct response (choice 1). (Reprinted with permission from Waltz et al., 1999.)

either the horizontal or vertical dimension; the other dimension was constant (figure 69.3*B*). Two-relation problems (level-2 complexity) required integrating two relational changes over the horizontal and vertical dimensions, respectively (figure 69.3*C*). Thus, although the basic form of the task was constant across the three types of matrix problems, only the two-relation problems necessitated relational integration.

Figure 69.4 presents the results obtained by Waltz and colleagues (1999) for the matrix problems. The temporal-variant FTLD patients and the normal controls achieved very high accuracy at all levels of relational complexity. The frontal-variant patients also performed at a high level for level-0 and level-1 problems; however, their performance plunged dramatically for level-2 items (just 11% correct, not different from chance). At the same time, the performance of frontal-variant patients was superior to that of temporal-variant patients on a test of recognition memory and on tests dependent on semantic knowledge. The resulting double dissociation between relational reasoning versus both episodic memory and semantic knowledge rules out a general "difficulty" factor as the source of the prefrontal group's impairment.

The frontal-variant patients also were completely unable to solve transitive-inference problems that required integration of two relations. Reasoning deficits of a similar but less pronounced nature have also been observed in patients with Alzheimer's disease with pronounced frontal signs (Waltz et al., 2004). Taken together, these findings indicate that the human PFC plays an essential role in relational reasoning—specifically, in the integration of multiple relations. Furthermore, the role of the PFC in relational integration

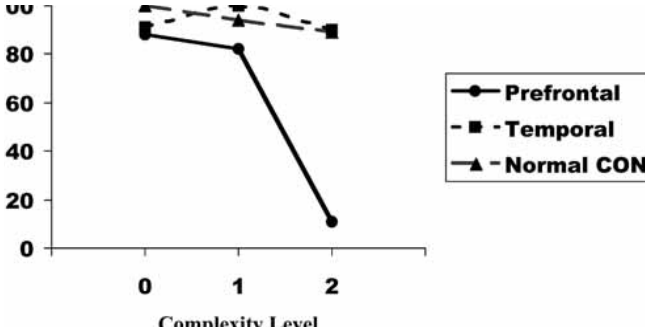


FIGURE 69.4 Accuracy on matrices test (Waltz et al., 1999). Groups show similar performance in the solution of nonrelational problems (level 0), as well as those requiring maintenance of a single relation (level 1); but patients with prefrontal damage show catastrophic impairment in the ability to solve problems requiring the integration of multiple relational premises (level 2). (Reprinted with permission from Waltz et al., 1999.)

was shown to generalize across both inductive (RPM) and deductive (transitive inference) reasoning tasks.

INTERFERENCE CONTROL AND THE PFC Other recent neuropsychological studies using FTLD patients have examined the role of the PFC in controlling interference from distracting information during analogical reasoning. Morrison and colleagues (2004) tested both frontal- and temporal-variant FTLD, as well as age-matched controls, on a verbal analogy task. Four-term analogy problems of the form A:B::C:D or D′ were employed, where D is the analogical answer and D′ is a nonanalogical foil (adapted from Sternberg & Nigro, 1980). A semantic facilitation index (SFI) was calculated for each problem to characterize the association of the correct relational pair (C:D) relative to the distracter pair (C:D′). For example, for the problem PLAY:GAME::GIVE: ? (1) PARTY (2) TAKE, the C:D pair (GIVE:PARTY, the correct analogical answer) is less associated than is the C:D′ pair (GIVE:TAKE, the nonanalogical foil), yielding a negative SFI for the problem. The problems were divided into those with negative SFI, neutral SFI, and positive SFI in order to examine the effect of semantic interference on the ability to identify the analogical answer.

Morrison and colleagues (2004) predicted that because these 4-term analogy problems were based on a single, fairly simple relation between the A and B terms, frontal-variant FTLD patients should be able to perform the basic analogical mapping needed despite their diminished working memory. However, if the PFC is also critical for interference resolution, then these patients should be selectively impaired on problems in which the D′ distracter is a strong competitor to the analogical choice, D. In the positive and neutral SFI conditions, the analogical answer (D) does not face competition from an alternative (D′) that is more strongly associated with the C term. Accordingly, the analogical answer can simply be activated and produced as a response. However, in the negative SFI condition the D′ foil is in fact more strongly associated with C than is the analogical response D. It follows that in order to make the analogical response for these problems, it will be necessary not just to activate the D response, but also to inhibit the semantically related D′ response.

Accordingly, because of their postulated deficits in inhibitory control, it was predicted that frontal-variant FTLD patients would be selectively impaired in the negative SFI condition relative to the positive and neutral SFI conditions. In contrast, it was predicted that temporal-variant FTLD patients would show a more uniform decline in verbal analogy performance across all three conditions because of their loss of the conceptual information necessary to encode the relations in the analogy problem. Both of these patterns were in fact observed by Morrison and colleagues (2004). A more recent study using 4-term picture analogies also found that frontal-variant FTLD patients are especially impaired on problems that include semantically related distracters (Krawczyk et al., 2008).

The deficits in the frontal- and temporal-lobe patient groups that Morrison and colleagues (2004) found with the verbal analogy task were modeled using LISA. It proved possible to simulate the observed pattern of frontal-lobe deficits by impairing the rate of rapid learning of analogical connections in LISA's working memory, coupled with reduction of a parameter for inhibitory control. Both the rapid learning of new connections in working memory and inhibitory control appear to be key functions of prefrontal cortex (Miller & Cohen, 2001; Shimamura, 2000). When both these functions (not either one alone) were impaired in LISA, the model yielded the selective impairment on negative SFI problems shown by frontal-lobe patients. Thus LISA would predict that distinct prefrontal regions would be activated during analogical reasoning corresponding to these two components. When the extent of semantic death (loss of connections between semantic units representing a relational role and the predicate unit for that role) was increased in LISA, thereby modeling loss of conceptual knowledge in anterior temporal cortex, the simulation yielded the pattern of impairment found for temporal-variant FTLD patients: a relatively uniform decrease in accuracy across all verbal analogy problems, regardless of SFI condition.

RELATIONAL REASONING IN THE AGING BRAIN The PFC has been shown to be vulnerable to the effects of aging (Raz et al., 1997). Decreased PFC function in older adults may affect the ability to cope with relational complexity and interference during reasoning. Viskontas and colleagues (2004) manipulated both factors using a set of analogy problems based on cartoons of human figures, each defined

by four binary-valued characteristics (height, girth, clothing color, and gender). Figure 69.5 illustrates examples of these People Pieces Analogies (PPA; Sternberg, 1977; Morrison, Holyoak, & Truong, 2001; Cho et al., 2007). In the PPA task, subjects are asked to verify as quickly as possible whether the relationship between the A:B pair corresponds to the relationship between the C:D pair with respect to selected dimension(s). This task has a major advantage over other paradigms because visual complexity is controlled by having all problems, regardless of the level of relational complexity, consist of four characters from the basic set of 16. Before each trial in the PPA task participants see a list of the traits that they are to attend to on that trial. They then see an analogy problem that they have to judge as true or false based on the dimensions specified as relevant. The number of traits to which participants attend can be varied from one to four, thus manipulating the relational complexity of the problem. In addition, 1- and 2-relation problems may have 0, 1, or 2 irrelevant dimensions that need to be inhibited to avoid interference with the analogical answer. "False" trials are constructed by introducing the wrong value on one relevant dimension.

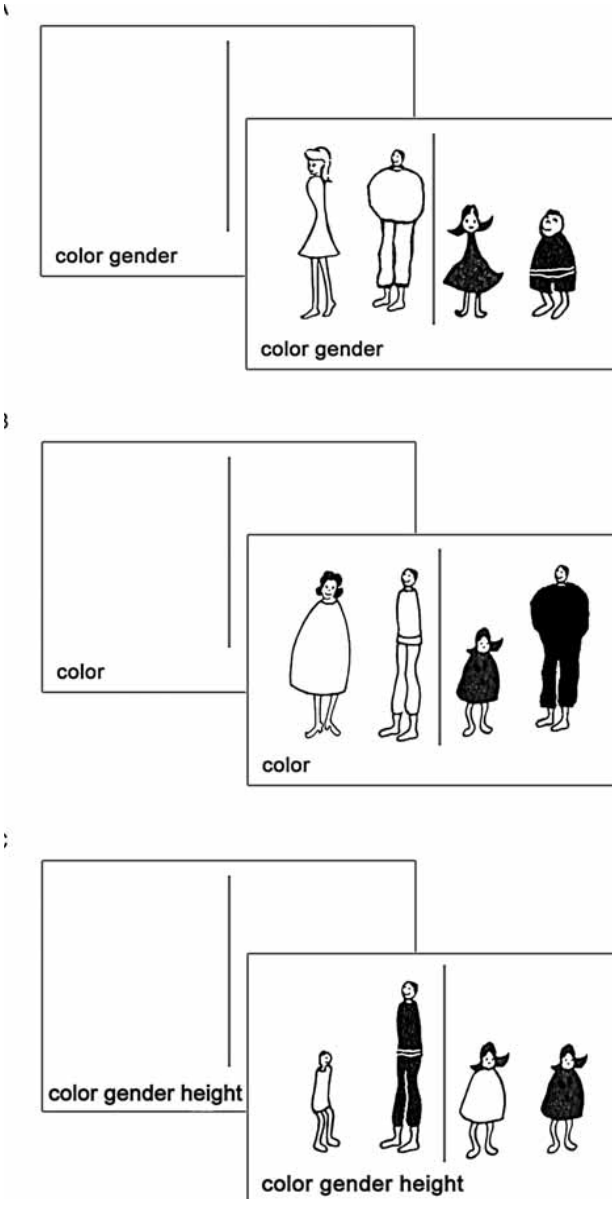


FIGURE 69.5 Example problems from People Pieces Analogy task (Viskontas et al., 2004). (*A*) two relations, none to inhibit; (*B*) one relation to attend to, two to inhibit; (*C*) problem for which correct response is "false." (Reprinted with permission from Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004.)

Viskontas and colleagues (2004) administered the PPA tasks to three groups of adult subjects: younger, middle-aged, and older (mean ages of approximately 20, 50, and 75 years, respectively). The mean response times (RTs) for correct "true" trials are shown in figure 69.6 (left). For both young and older subjects, RTs were significantly increased for greater levels of relational complexity (compare figure 69.6*A* to 69.6*C*). For young subjects RTs increased only slightly when distracting information was present. In older adults, however, distracting information had a profound effect on response time when it was necessary to integrate more than one relation (figure 69.6*C*). Age-related impairments in relational integration have also been obtained with other deductive and inductive reasoning tasks (Viskontas et al., 2005).

Viskontas and colleagues (2004) showed that the LISA model could successfully capture the interaction among age, complexity, and the amount of distracting information simply by varying an inhibition parameter. Reducing the model's inhibition parameter has the effect of increasing sensitivity to distracting information, particularly when multiple relations need to be mapped in working memory (see figure 69.6, right). Such loss of inhibitory control for older adults is consistent with the decline in prefrontal functions that accompanies cognitive aging. In younger adults, performance on analogy problems that require interference control can be impaired by imposing dual-task conditions (Morrison et al., 2001; Waltz, Lau, Grewal, & Holyoak, 2000), by delaying the cue signaling the relevant dimensions on a trial (Cho et al., 2007), or by inducing anxiety prior to administering the analogy task (Tohill & Holyoak, 2000). All these factors plausibly act by imposing excessive load on working-memory processes dependent on the PFC.

Functional imaging of component processes in relational reasoning

Neuropsychological evidence has established the critical role of the PFC in relational reasoning, and behavioral and computational modeling studies point to the distinctive roles of relational integration and inhibitory control. In recent years, considerable progress has been made in localizing specific

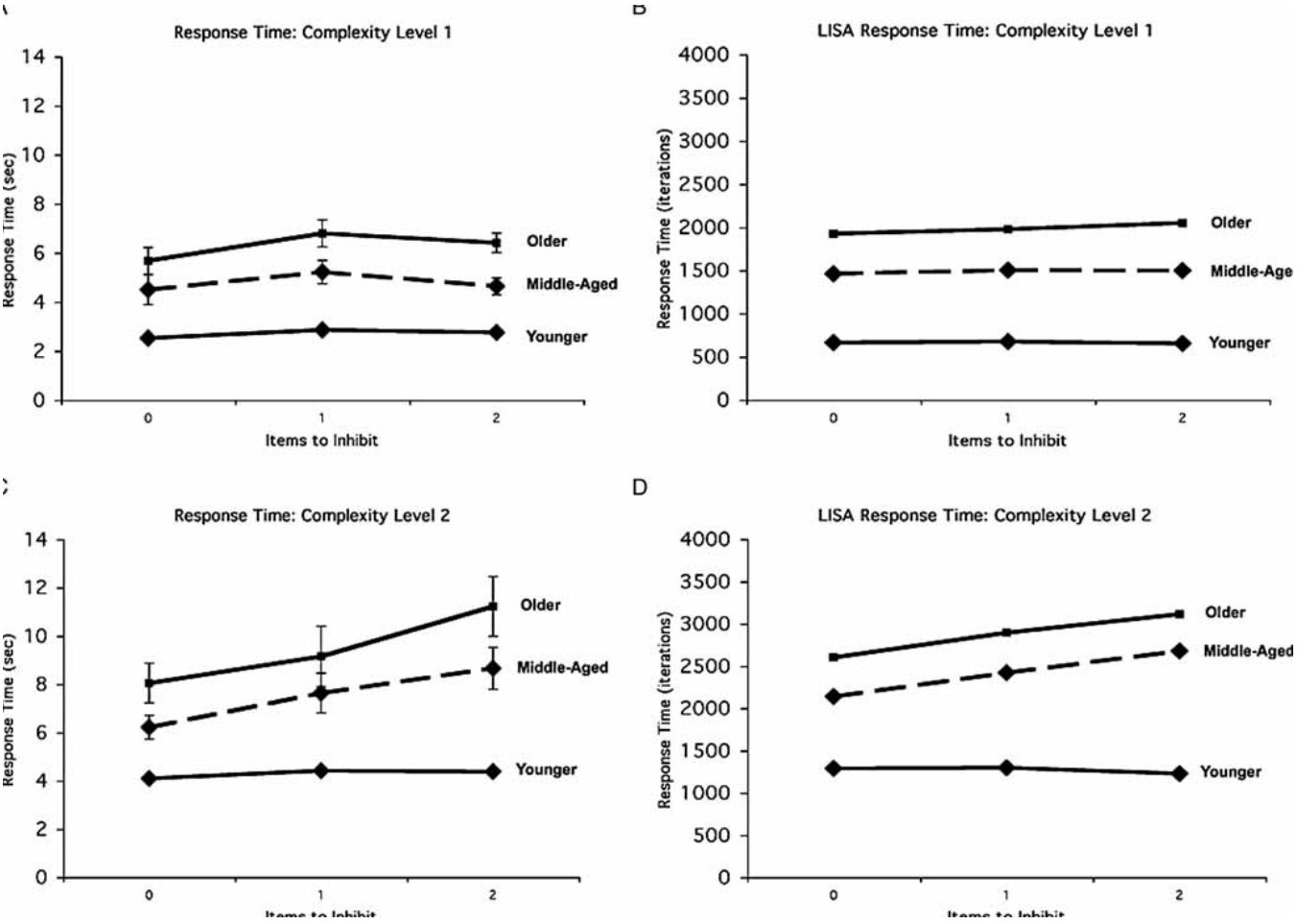


FIGURE 69.6 (*A*) Response time in the People Pieces Analogy task for 3 levels of inhibition at complexity level 1 for younger, middle-aged, and older groups; (*B*) corresponding LISA simulations; (*C*) human data for complexity level 2; and (*D*) corresponding LISA simulations. Error bars depict standard error of the mean. (Reprinted with permission from Viskontas et al., 2004.)

PFC and other brain areas associated with relational reasoning using PET (Wharton et al., 2000) and fMRI.

RELATIONAL INTEGRATION Several fMRI studies have manipulated relational complexity using variants of RPM problems, similar to those used by Waltz and colleagues (1999) with FTLD patients. With healthy young adults, more complex problems can be used (typically 3×3 matrices with up to four dimensions of variation; Kroger et al., 2002). For matrix problems, relational integration has been shown to consistently activate prefrontal regions. In particular, bilateral middle (MFG) and inferior (IFG) frontal gyri, as well as parietal and occipital regions, have been found to increase activity when multiple relations must be integrated in order to arrive at a solution, compared to problems that require processing of only a single relation (Prabhakaran et al., 1997; Christoff et al., 2001; Kroger et al., 2002).

Among these regions, which constitute a network commonly activated in visuospatial working memory tasks, the activation pattern of the most anterior part of the PFC has been particularly noteworthy. Christoff and colleagues (2001) found that the left frontopolar region remained preferentially activated even after controlling for the influence of increased problem-solving time. Kroger and colleagues (2002) confirmed and extended these results, providing evidence that although the left anterior prefrontal region becomes increasingly activated as more relations need to be integrated, activation in this subarea is not affected by increases in perceptual difficulty (achieved by adding visuospatial distracters to RPM problems while holding relational complexity constant). Thus the frontopolar region seems to be uniquely associated with the specific requirement of integrating multiple relations, and not general cognitive difficulty.

Similarly, studies of verbal analogical reasoning have distinguished neural substrates of reasoning from semantic processing demands within working memory. Activation in the left frontopolar region increases selectively when making

judgments of analogical similarity compared to processing of semantic associations or categories (Bunge et al., 2005; Green et al., 2006; Wendelken et al., 2008). Thus, based on a substantial body of findings involving solution of different types of relational reasoning problems, the frontopolar region seems to play a special role in the process of integrating multiple relational representations to arrive at a solution.

Separating Relational Complexity and Interference Control While several studies have examined the neural correlates of relational integration, less is known about the neural basis of interference control in the context of analogical reasoning. Interference resolution (often linked to selection and inhibitory control) has been extensively studied using a wide variety of paradigms. Studies have identified the lateral PFC, including regions in the dorsolateral PFC and inferior frontal gyrus, as important for interference resolution across diverse tasks including inhibition of a motor response, proactive interference resolution in working memory, selection among competing alternatives, and controlled semantic retrieval (Aron, Robbins, & Poldrack, 2004; Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; D'Esposito, Postle, Jonides, & Smith, 1999; Jonides, Smith, Marshuetz, Koeppe, & Reuter-Lorenz, 1998; Nee, Wager, & Jonides, 2007; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Thompson-Schill et al., 2002; Wagner, Pare-Blagoev, Clark, & Poldrack, 2001; Zhang, Feng, Fox, Gao, & Tan, 2004). It seems plausible that these regions also support interference resolution in relational reasoning.

Recently, work in our lab (Cho et al., 2009) has jointly examined the neural substrate for relational integration and for interference resolution using the PPA task described earlier (Viskontas et al., 2004). In this task, it is possible to vary both factors simultaneously while holding visuospatial complexity constant. As in previous studies using the PPA task, the subject had to determine whether the analogy between the two pairs was valid, based on a subset of trait(s) randomly selected for each trial. Participants were instructed to solve each problem based on relevant ("to-be-attended") trait(s) only, to ignore irrelevant ("to-be-ignored") traits, and to decide as quickly and accurately as possible. A trait list consisting of four words naming each trait was displayed in the center of the screen, between the A:B and C:D pairs of cartoon characters. On each 8-second trial, the target pair (A:B) appeared on the screen for 1.7 s (target phase). The trait names were all shown in black font during the target phase. After the target pair disappeared, the to-be-attended trait cue(s) turned red (cue phase) and remained on the screen. After 0.3 s, the probe pair (C:D) appeared on the right side of the screen (probe phase), and subjects were allowed a maximum of 6 s to respond with a key press. The delayed cuing of to-be-attended traits ensured that subjects had to actively pay attention to all visual information about the A:B pair and that potential sources of interference would therefore be encoded into working memory.

Relational complexity level (number of to-be-attended traits, 1 or 3) and need for interference resolution (number of to-be-ignored traits that supported an incorrect response, 0 or 1) were manipulated in a factorial event-related fMRI design. There was a large increase in response time (RT) for PPA problems at the higher level of relational complexity. There was also an interactive effect of interference, which resulted in a reliable increase in RT only at the higher level of relational complexity. This overadditive interaction in the RT data resembled that found in previous studies using the PPA task (Cho et al., 2007; Viskontas et al., 2004). Both higher complexity and presence of interference significantly reduced accuracy in solving the analogy problems.

The fMRI analyses revealed cortical regions sensitive to increase in demands on relational integration, interference resolution, or both component processes of reasoning (see figure 69.7). In a whole-brain analysis, clusters activated by an increase in relational complexity were identified in bilateral frontal pole, as well as other regions including the MFG and IFG. In an a priori defined anatomical region-of-interest analysis of the lateral PFC comprising bilateral MFG and IFG, regions sensitive to increase in demands on interference resolution were found in bilateral MFG and IFG pars opercularis and the IFG pars triangularis in the right hemisphere. Regions sensitive to both component processes of analogical reasoning were found in bilateral MFG and IFG pars opercularis and the right IFG pars triangularis. It is possible that the neural basis for the behavioral RT interaction that was observed between relational complexity and interference may lie in this region of spatial overlap of activation. As suggested by the LISA model, inhibitory control is particularly critical when multiple relations are mapped in working memory. These overlapping regions may thus be involved in inhibitory control in working memory.

Consistent with proposals that the frontopolar region plays a specific role in reasoning when the outcomes of two or more relational comparisons must be integrated to arrive at a solution (Waltz et al., 1999; Christoff et al., 2001; Kroger et al., 2002), Cho and colleagues (2009) found signal increases in bilateral frontal pole related to increases in relational complexity but not to increased difficulty created by adding interference. By eliminating alternative explanations based on visual complexity or general cognitive difficulty, these results provide strong evidence supporting the hypothesis that the frontal pole plays a distinct role in reasoning tasks that require the integration of multiple relations.

The region-of-interest analyses found that areas of the lateral PFC that have been identified as important in executive control in a variety of cognitive tasks (dorsolateral and inferior frontal region) are also activated by the need

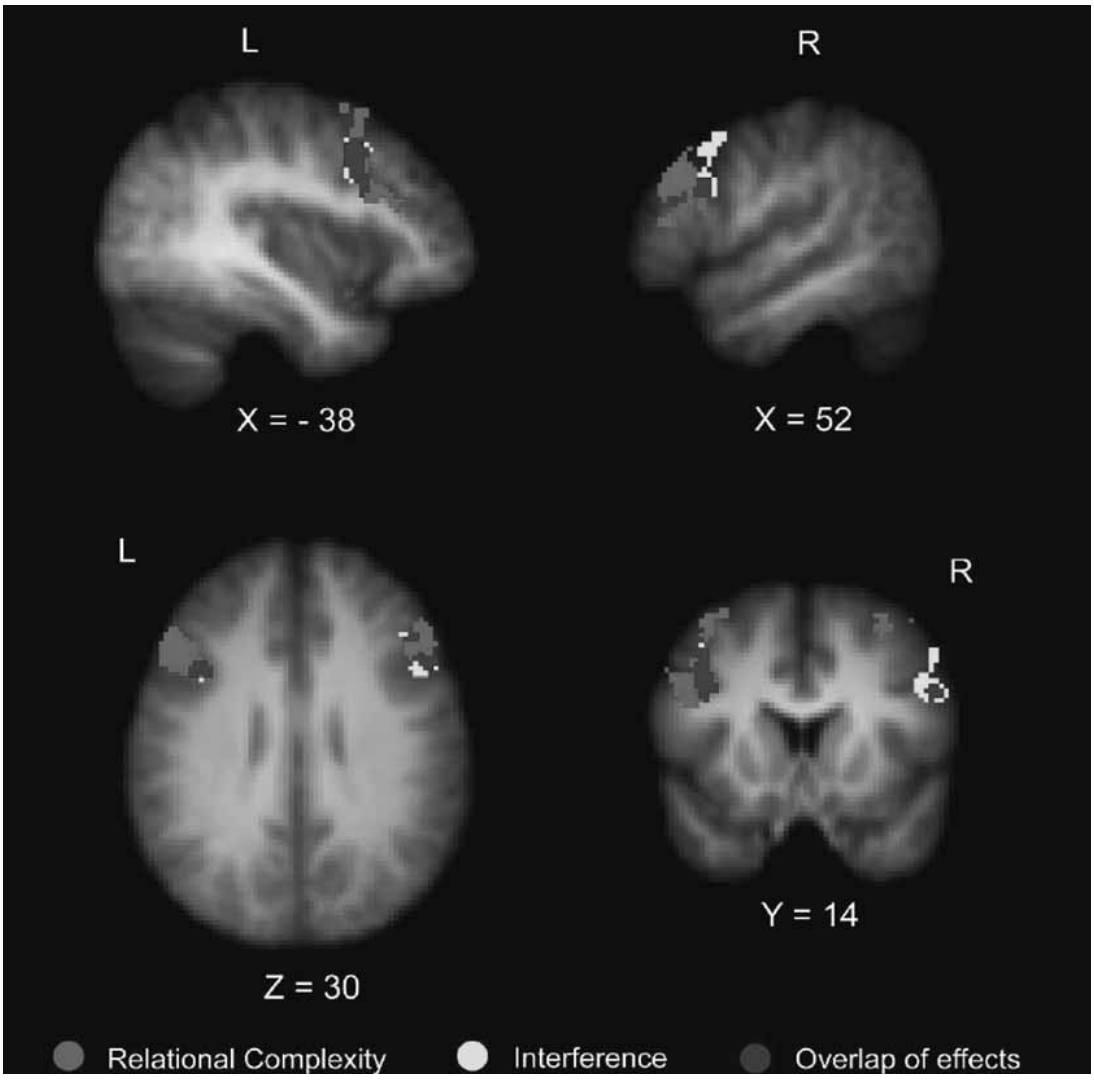


FIGURE 69.7 Neuroimaging results from Cho and colleagues (2009). Regions showing the main effects of relational complexity (shown in red), interference (shown in yellow; small volume corrected, uncorrected cluster-forming threshold $T > 2.3$, corrected cluster extent significance threshold, $p < .05$), and regions where main effects overlapped (blue) within an a priori defined anatomical ROI mask of the bilateral MFG and IFG pars opercularis and pars triangularis. R, right; L, left. Coordinates are in MNI space (mm). (See color plate 83.)

to resolve interference during analogical reasoning. This finding is consistent with a broad role for these regions in cognitive control. By manipulating multiple cognitive demands simultaneously in a single reasoning task, the Cho and colleagues (2009) study was able to elucidate the neural architecture underlying behavioral interactions between complex cognitive processes. These findings show that analogical reasoning, which requires integration of multiple relations in the face of interference, is associated with the coordination of activity in multiple functionally dissociable regions of the prefrontal cortex. These subregions include those that are relatively more sensitive to demands on one component process, as well as regions that are jointly taxed by both relational integration and interference resolution.

THE TOPOGRAPHY OF THINKING As our survey of recent research makes clear, recent neuropsychological studies have demonstrated the dependence of relational reasoning on PFC, and neuroimaging studies have begun to delineate the functional anatomy of reasoning processes. These studies have converged on the finding that the frontopolar region is specifically activated when a problem requires the integration of multiple relations. As reviewed by Ramnani and Owen (2004), this region has several properties consistent with its playing an important role in complex cognitive tasks. First, the frontopolar region is comparatively larger in humans than in nonhuman primates. Second, unlike other PFC regions, it primarily has reciprocal connections with other supramodal regions in the PFC, suggesting that it is suited for processing abstract information. Third, the cellular

properties of the frontopolar cortex are consistent with its being a site of convergence. Although the density of neurons is not particularly high in this region, postsynaptic spines are extremely plentiful and dense in the dendritic arbors of these neurons.

While neuroimaging data from studies using reasoning tasks is consistent with the hypothesis that the frontopolar cortex subserves relational integration, this region is also activated in several other tasks involving complex cognition. For example, frontopolar cortex is active when subjects interrupt one task temporarily to perform another (a task requirement termed "branching"; Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999) and during episodic memory retrieval (Rugg et al.,1998). It may be possible to develop a theory of frontopolar cortex function that can account for the full range of findings. According to one proposal, the role of the frontopolar cortex is to process internally generated information (Christoff & Gabrieli, 2000). Relational integration might be subserved under this view if it is assumed that the reasoner must manipulate internally generated relations. On the face of it, however, the basic relations that must be integrated in a task such as the RPM (e.g., change in size or shape of a figure) are perceptually available in the problem itself. According to another view, the role of the frontopolar cortex is to integrate the results of multiple cognitive operations (Ramnani & Owen, 2004). Rather than focusing on the type of information being manipulated, this view emphasizes the cognitive processes supported by this region. Ramnani and Owen's proposal appears to be consistent with the relational integration hypothesis, as the outputs of cognitive processes underlying planning and problem solving (e.g., selected operators and new subgoals) may constitute relations that need to be integrated in order to sequence actions.

Another approach to developing a more precise theory of frontopolar functions is to search for subdivisions that support distinct cognitive processes. Recently, Wendelken and colleagues (2008) argued that the frontopolar region activated by relational integration in analogical reasoning is more lateral than the frontopolar region activated by branching, which is in the most rostral aspect of this region. A similar conclusion was reached by Gilbert and colleagues (2006) in a meta-analysis of neuroimaging studies showing frontopolar activations. Gilbert and colleagues also concluded that multitasking paradigms elicited more rostral activation. Thus it is possible that the most rostral subregion is involved in coordinating multiple task sets and goals, whereas more caudal, lateral subregions are involved in integration of relations (or perhaps, those relations represented in an explicit, declarative code of the sort employed in models of relational thought, such as LISA).

It also appears that there may be dorsoventral differences within the lateral frontopolar cortex, with more dorsal areas activated during solution of visuospatial problems such as the RPM task and more ventral areas activated during verbal analogy tasks. Such variations may reflect a difference in the type of materials used in the tasks (visuospatial versus verbal), or other differences between the tasks (e.g., generating an analogical solution versus evaluating an analogical mapping). Future research using high-resolution fMRI may be able to further tease apart the contributions of distinct subregions of frontopolar cortex.

Over the past decade, understanding the neural basis of relational reasoning has for the first time become a tractable research problem. Much of the recent progress has been the product of neuroimaging and neuropsychological studies guided by theoretical frameworks developed in cognitive science. As the component processes of reasoning are mapped onto brain regions, it will be possible to delineate a functional network for reasoning. This accomplishment will represent a major step toward understanding the remarkable capabilities of human thought.

ACKNOWLEDGMENTS This chapter was written with support from the Office of Naval Research (Grant N000140810186).

REFERENCES

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.*, *8*, 170–177.

Asaad, W. F., Rainer, G., & Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron*, *21*, 1399–1407.

Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*, 907–918.

Brun, A. (1993). Frontal lobe degeneration of non-Alzheimer's type revisited. *Dementia*, *4*, 126–131.

Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cereb. Cortex*, *15*, 239–249.

Cho, S., Holyoak, K. J., & Cannon, T. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Mem. Cogn.*, *35*, 1445–1455.

Cho, S., Moody, T. D., Fernandino, L., Mumford, J. A., Poldrack, R. A., Cannon, T. D., Knowlton, B. J., & Holyoak, K. J. (2009). Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. Manuscript under review.

Christoff, K., & Gabrieli, J. D. E. (2000). The frontopolar cortex and human cognition: Evidence for a rostro-caudal hierarchical organization within the human prefrontal cortex. *Psychobiology*, *28*, 168–186.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, *14*, 1136–1149.

D'Esposito, M., Postle, B. R., Jonides, J., & Smith, E. E. (1999). The neural substrate and temporal dynamics of interference

effects in working memory as revealed by event-related functional MRI. *Proc. Natl. Acad. Sci. USA*, *96*, 7514–7519.

Doumas, L. A. A., & Hummel, J. E. (2005). Approaches to modeling human mental representations: What works, what doesn't, and why. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 73–91). Cambridge, UK: Cambridge University Press.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artif. Intell.*, *41*, 1–63.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cogn. Sci.*, *7*, 155–170.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cogn. Psychol.*, *12*, 306–355.

Giedd, J. N. (2004). Structural magnetic resonance imaging of the adolescent brain. *Ann. NY Acad. Sci.*, *1021*, 105–109.

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., & Burgess, P. W. (2006). Functional specialization within rostral prefrontal cortex (area 10): A meta-analysis. *J. Cogn. Neurosci.*, *18*, 932–948.

Green, A. E., Fugelsang, J. A., Kraemer, D. J., Shamosh, N. A., & Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Res.*, *1096*, 125–137.

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Erlbaum.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behav. Brain Sci.*, *21*, 803–831; discussion, 831–864.

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1996). Semantic dementia: Progressive fluent aphasia with temporal lobe atropy. *Brain*, *115*, 1783–1806.

Holyoak, K. J. (2005). Analogy. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 117–142). Cambridge, UK: Cambridge University Press.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cogn. Sci.*, *13*, 295–355.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychol. Rev.*, *104*, 427–466.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychol. Rev.*, *110*, 220–264.

Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., & Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proc. Natl. Acad. Sci. USA*, *95*, 8410–8413.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, *399*, 148–151.

Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., Miller, B. L., & Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, *46*, 2020–2032.

Kroger, J. K., Saab, F., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cereb. Cortex*, *12*, 477–485.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.*, *24*, 167–202.

Morrison, R. G., Doumas, L. A. A., & Richland, L. E. (2006). The development of analogical reasoning in children: A computational account. In R. Sun & N. Miyake (Eds.), *Proceedings of the Twenty-eighth Annual Conference of the Cognitive Science Society* (pp. 603–608). Mahwah, NJ: Erlbaum.

Morrison, R. G., Holyoak, K. J., & Truong, B. (2001). Working memory modularity in analogical reasoning. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society* (pp. 663–668). Mahwah, NJ: Erlbaum.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *J. Cogn. Neurosci.*, *16*, 260–271.

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cogn. Affect. Behav. Neurosci.*, *7*, 1–17.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.*, *31*, 109–178.

Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cogn. Psychol.*, *33*, 43–63.

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights from anatomy and neuroimaging. *Nat. Rev. Neurosci.*, *5*, 184–194.

Raven, J. C. (1941). Standardization of progressive matrices, 1938. *Br. J. Med. Psychol.*, *19*, 137–150.

Raz, N., Gunning, F. M., Head, D., Dupuis, J. H., McQuain, J., Briggs, S. D., Loken, W. J., Thornton, A. E., & Acker, J. D. (1997). Selective aging of the human cerebral cortex observed in vivo: Differential vulnerability of the prefrontal gray matter. *Cereb. Cortex*, *7*, 268–282.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *J. Exp. Child Psychol.*, *94*, 249–271.

Robin, N., & Holyoak, K. J. (1995). Relational complexity and the functions of prefrontal cortex. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 987–997). Cambridge, MA: MIT Press.

Rugg, M. D., Fletcher, P. C., Allan, K., Frith, C. D., Frakowiak, R. S., & Dolan, R. J. (1998). Neural correlates of memory retrieval during recognition memory and cued recall. *NeuroImage*, *8*, 262–273.

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology*, *28*, 207–218.

Snowden, J., Neary, D., & Mann, D. (2007). Frontotemporal lobar degeneration: Clinical and pathological relationships. *Acta Neuropathol.*, *114*, 31–38.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.

Sternberg, R. J., & Nigro, G. (1980). Developmental patterns in the solution of verbal analogies. *Child Dev.*, *5*, 127–138.

Thagard, P. (1996). The concept of disease: Structure and change. *Commun. Cogn.*, *29*, 445–478.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc. Natl. Acad. Sci. USA*, *94*, 14792–14797.

Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D'Esposito, M., Kan, I. P., Knight, R. T., & Swick, D. (2002).

Effects of frontal lobe damage on interference effects in working memory. *Cogn. Affect. Behav. Neurosci.*, *2*, 109–120.

Tohill, J. M., & Holyoak, K. J. (2000). The impact of anxiety on analogical reasoning. *Thinking & Reasoning*, *6*, 27–40.

Viskontas, I. V., Holyoak, K. J., & Knowlton, B. J. (2005). Relational integration in older adults. *Thinking & Reasoning*, *11*, 390–410.

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition and analogical reasoning in older adults. *Psychol. Aging*, *19*, 581–591.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*, 329–338.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Back-Madruga, C., McPherson, S., Masterman, D., Chow, T., Cummings, J. L., & Miller, B. L. (2004). Relational integration and executive function in Alzheimer's disease. *Neuropsychology*, *18*, 296–305.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., Santos, M. D., Thomas, C. R., & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychol. Sci.*, *10*, 119–125.

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Mem. Cogn.*, *28*, 1205–1212.

Wendelken, C., Nakahbenko, D., Donohue, S. E., Carter, C. S., & Bunge, S. A. (2008). "Brain is to thought as stomach is to ??" Investigating the role of rostrolateral prefrontal cortex in relational reasoning. *J. Cogn. Neurosci.*, *20*, 682–693.

Wharton, C. M., Grafman, J., Flitman, S. S., Hansen, E. K., Brauner, J., Marks, A., & Honda, M. (2000). Toward neuroanatomical models of analogy: A positron emission tomography study of analogical mapping. *Cogn. Psychol.*, *40*, 173–197.

Zhang, J. X., Feng, C. M., Fox, P. T., Gao, J. H., & Tan, L. H. (2004). Is left inferior frontal gyrus a general mechanism for selection? *NeuroImage*, *23*, 596–603.

70 Decision Making and Prefrontal Executive Function

CHRISTOPHER SUMMERFIELD AND ETIENNE KOECHLIN

ABSTRACT Decision making—the selection of one course of action from among many—critically involves the frontal lobes. Here, we propose a theory of the functional architecture of the human frontal lobes that details the cognitive and motivational processes underlying decision making. We argue for a fractionation of both lateral and medial prefrontal cortices into hierarchical, rostrocaudally arranged control processes, permitting action selection to be guided by task-related information arising across multiple contexts and episodes, and in accordance with likely investment and outcome over both the short and long term. We suggest that the temporal dimension of the decision process may be the key to understanding the functional organization of the frontal lobes, with caudal frontal regions processing the cognitive and motivational demands of immediate action selection, and more rostral frontal cortex responsible for maintaining a task set and its associated value across an extended behavioral episode. Finally, we discuss how cognitive and motivational information might be integrated into a central executive system in the service of decision making.

Decision making is the cognitive process that allows one course of action to be chosen from among several possible options. The human brain is capable of making decisions of astonishing complexity—for example, when a great number of hypothetical outcomes are compared in order to select the winning move in a game of chess. Much of the time, however, the brain is occupied with much more mundane choices—for example, should I reach for that opportunely placed cup of coffee and take a sip? Even quotidian decisions such as this can require a great wealth of information to be integrated in order for the optimal course of action to be identified. First, on a visceral or affective level, many decisions depend on personal preferences—a good starting point might be simply whether you like the taste of coffee. Second, in the vast majority of decisions, a range of complex motivational factors will come into play. What are the likely costs and benefits associated with drinking that cup of coffee—even if it is warm and tasty, might it be too strong, and make me feel rather jittery and nervous? Finally, these motivational factors interact at each stage with a cognitive architecture that brings relevant contextual information from past and current episodes to bear upon the decision. For example, if the coffee is sitting on the desk in front of you, but belongs to your boss, you might think twice about reaching for it and taking a sip.

How then, does the brain integrate all this cognitive and motivational information to make the right choice? After many years, a picture is emerging of how decision making is controlled by the frontal lobes, that is, the anterior portion of the cerebral cortex stretching from the motor cortex to the rostral pole (see figure 70.1). In this review, we argue that discrete sectors of the frontal lobes make distinct contributions to decision making, with the orbital sector underlying the affective value of the stimulus (e.g., I like coffee), the medial sector controlling the motivation to act (I want coffee), and the lateral sector overseeing the higher-order cognitive structure of plans or goals (I *select* a coffee).

Second, we argue for *hierarchical control* over decision making. Some decisions—such as selecting to press the brake pedal (rather than the accelerator) at a red traffic light—can be undertaken automatically, on the basis of overlearned, habitual routines that require little intervention from the executive processes typically associated with the brain's most anterior zones. Other decisions—for example, electing to drive through a red traffic light, once you suspect that it is malfunctioning—require these sensorimotor affordances to be integrated with specific information pertaining to the current context—for example, the fact that the wait has been unusually long or that other drivers have also chosen to ignore the red light. We call this type of decision process, where basic sensorimotor codes are integrated with information from the current situation, *contextual control*. Finally, relevant past experience or recent task instructions can add yet another layer of pertinent information to a decision. If you have been waiting in front of a red traffic light for several minutes, but you have also just driven past a sign warning that long delays are anticipated, that past information can help you decide whether to continue your frustrating wait. We call this process, by which yet more temporally distant information is brought to bear upon a decision, *episodic control*, because the instruction cue defines the onset or offset of a discrete episode in which a given set of behavioral rules apply. We will argue that sensorimotor, contextual, and

CHRISTOPHER SUMMERFIELD Oxford University, Oxford, United Kingdom

ETIENNE KOECHLIN Institut National de la Santé et de la Recherche Médicale; Université Pierre et Marie Curie; Ecole Normale Supérieure, Paris, France

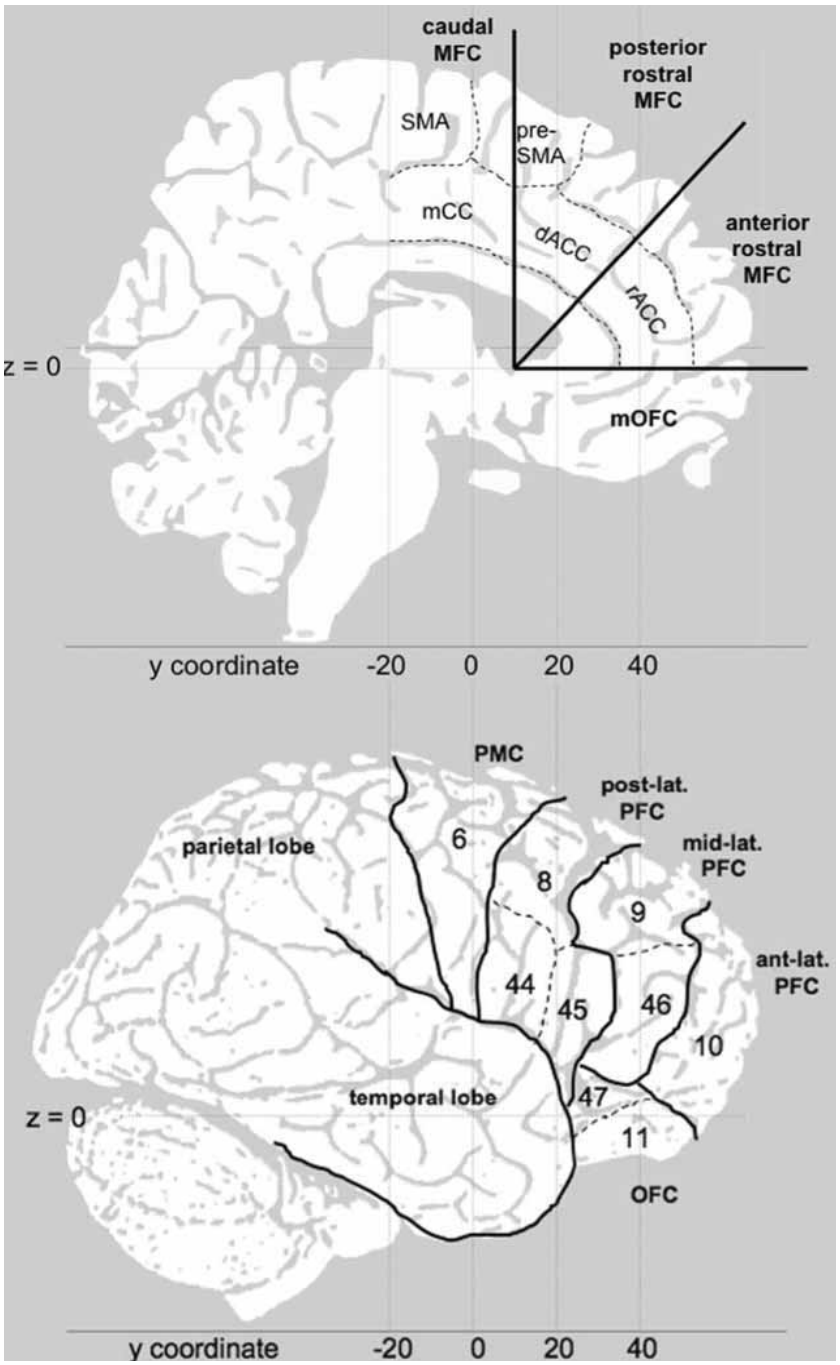


FIGURE 70.1 Anatomy of the human frontal lobes. Medial (top) and lateral (bottom) views of the frontal lobes in the Talairach coordinate stereotaxic system. Numbers indicate Brodmann's areas. PMC, premotor cortex; MFC, medial frontal cortex; OFC, orbitofrontal cortex; PFC, prefrontal cortex; SMA, supplementary motor area; ACC, anterior cingulate cortex; mCC, midcingulate cortex.

episodic control have both a cognitive dimension, pertinent to the selection of the appropriate action, and a motivational dimension, relevant to calculating the likely investment or outcome associated with a decision (for brevity, we do not discuss the orbitofrontal cortex and the affective component of decision making). Moreover, we will argue that the functional organization of the frontal cortex reflects this dissociation, with the neural substrates of episodic, contextual, and sensorimotor cognitive control implemented in a hierarchical, anteroposterior fashion along its lateral surface. Last, we will present some new evidence that the medial frontal cortex may exhibit a parallel, hierarchical organization for the motivational dimension of episodic and contextual information.

The lateral frontal cortex: Selection of responses, tasks, and goals

Two Views of Lateral Frontal Function Historically, our understanding of the function of the lateral frontal cortex (LFC) has been informed by two distinct strands of research. The first originated with now classic electrophysiological recordings in the awake behaving monkey, which demonstrated that prefrontal neurons exhibit tonic spiking activity across a delay period in which information has to be temporarily stored (Fuster & Alexander, 1971). Extensive neuroimaging work has built upon this finding to reveal that the LFC is engaged when humans perform short-term memory tasks, analogous to maintaining a telephone number for a span of a few seconds (Curtis & D'Esposito, 2003). The second major framework for characterizing research into the PFC argues that the lateral prefrontal cortex is the source of "biasing" signals that flow back to more posterior cerebral zones. These signals can selectively modulate perceptual (Desimone & Duncan, 1995) or motor (Passingham, 1996) codes on the basis of the currently active task or goal (Miller & Cohen, 2001). A common theme running through both these accounts, thus, is a role for the LFC in selecting actions on the basis of stimuli that are absent, obscured by distracters, or too novel or unexpected to benefit from automated behavioral routines formed by habit (Shallice, 1988). This function has been termed *cognitive control*, and it has proved a useful concept for summarizing the global significance of the prefrontal cortex in orchestrating thought and behavior (Miller & Cohen, 2001). However, despite the notable success of these two perspectives in explaining a wealth of data from brain imaging and electrophysiological recordings, neither view provides an entirely satisfactory answer to the twin problems of functional specialization (how the LFC is organized) and functional integration (how information flows through the LFC) during decision making (Koechlin & Summerfield, 2007).

Hierarchical Control of Action Selection in the LFC As outlined in the preceding paragraph, action selection can be conceived as a hierarchical process, in which both past and present contextual information can be integrated into the decision process along with elementary associative links between stimulus and response. A hierarchical structure to the prefrontal cortex has been proposed in which successively more anterior prefrontal regions encode increasingly abstract "perception-action" complexes—sustained neural activations that span the bridge between input and output—by way of reciprocal connectivity with parallel hierarchies in the posterior neocortices (Fuster, 2000). Indeed, this view squares well with the widespread observation that content- or domain-specific representations (for example, spatial versus object codes) tend to be more posteriorly in the LFC (Brodmann's area [BA] 8/44), whereas mid- and anterolateral prefrontal regions are less sensitive to the contents of the information being treated, but rather encode more general aspects of task structure (Christoff & Gabrieli, 2000; Ramnani & Owen, 2004; Sakai & Passingham, 2003). Indeed, the view that the frontal cortex comprises a pyramid for action selection, with abstract plans and goals encoded in more anterior regions, and concrete motor acts in motor and premotor cortical zones, is the centerpiece of several recent theories of lateral prefrontal function (Badre & D'Esposito, 2007; Fuster, 2000; Koechlin & Summerfield, 2007; Petrides, 2005).

The Cascade Model Koechlin and colleagues (Koechlin, Ody, & Kouneiher, 2003; Koechlin & Summerfield, 2007) have proposed a hierarchical account of decision making in the LFC that suggests a tripartite division among the biasing mechanisms that underpin top-down control of action (figure 70.2). Deciding which action to select, it is argued, requires the convergence of information from three different types of control process: "episodic" control, the tonic maintenance of task-relevant information across the interval separating an instruction cue from the decision; "contextual" control, reflecting the need to integrate information from contextual cues in the immediate environment; and "sensorimotor" control, treating the association between a stimulus and response. Koechlin and colleagues suggest that episodic, contextual, and sensorimotor control are implemented in an anteroposterior fashion along the LFC, with midlateral prefrontal regions (BA 9/46) underpinning episodic control, posterior lateral PFC regions (BA 8/44) subserving contextual control, and sensorimotor control the province of the premotor cortex (lateral BA 6). At the heart of the cascade model is the notion of *subsidiarity*, whereby successively higher regions are recruited only when processing in lower regions is not sufficient to allow optimal decision making to occur. For example, sensorimotor control (the selection of an action on the basis of a simple stimulus-response routine) requires only the premotor cortex but not higher regions, whereas contextual control (where the response is contingent both on the stimulus and context) requires posterior lateral PFC *and* premotor regions to come online. By extension, episodic control requires all three nodes in the chain to be active. Neural information thus "cascades" down the action selection hierarchy, from anterior to posterior lateral frontal regions, to be integrated at the base, in premotor cortex, where a response can then be prepared and passed to primary motor zones for execution. This model thus describes the gross functional organization of the LFC and simultaneously outlines how information flows between lateral prefrontal regions during decision making—thereby offering a joint solution to the problems of function

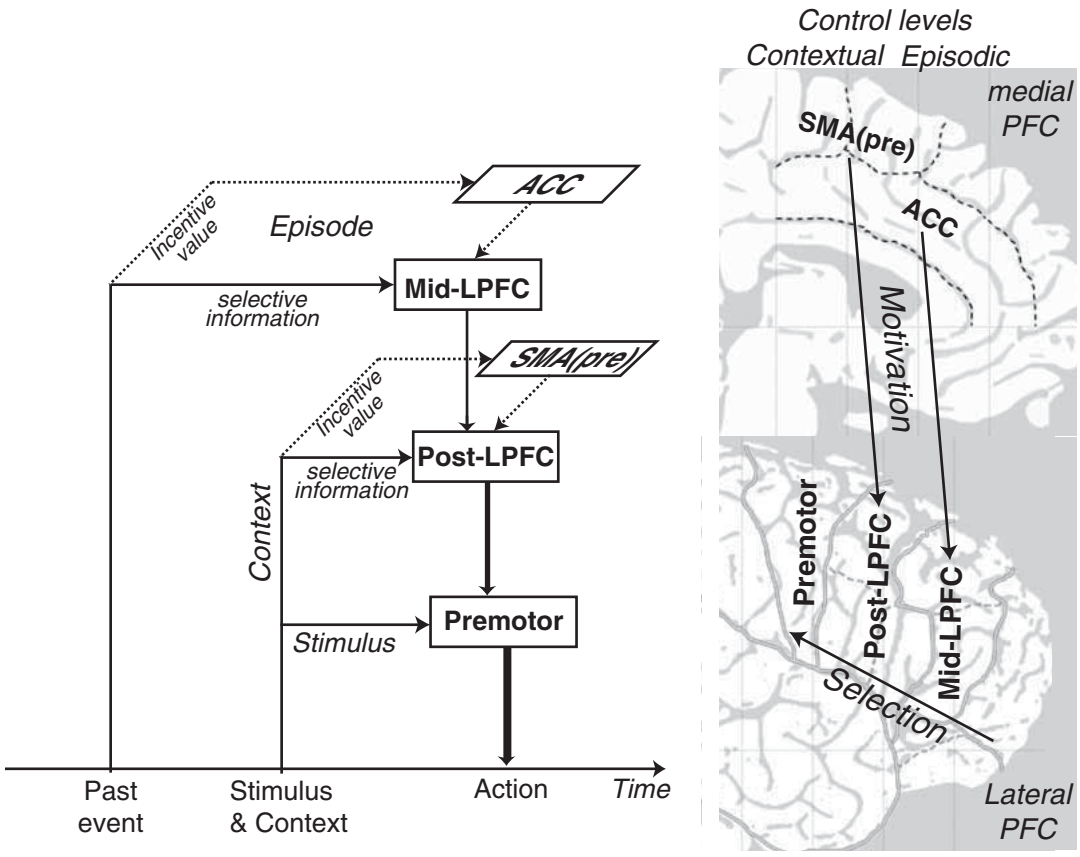


FIGURE 70.2 A model of the architecture of prefrontal executive function. The posterior and middle PFC subserve contextual and episodic control of behavior, respectively. The lateral PFC is involved in contextual and episodic selection based on information extracted from the immediate context and occurrence of past events, respectively. Selection operates through top-down interactions from anterior to posterior lateral PFC regions. The medial PFC is involved in contextual and episodic motivation based on incentive values extracted from the immediate context and occurrences of past events. Motivation operates through lateral interactions between medial and lateral PFC regions.

localization and functional integration in the frontal cortex (Koechlin and Summerfield, 2007).

EVIDENCE FOR THE CASCADE MODEL Despite its apparent simplicity, this "cascade" model can account for a wide range of data obtained from neurocognitive paradigms directed at understanding the function of the frontal lobes.

Premotor cortex The cascade model noncontroversially supports a rich literature implicating the premotor cortex in basic stimulus-response selection. For example, classic lesion experiments in the monkey demonstrated that selecting a movement in response to a colored cue—with or without an interposed delay—is greatly impaired following bilateral removal of the premotor cortex (Passingham, 1988). Single-cell recordings have confirmed that premotor neurons integrate information from a visual (or somatosensory) cue and its relevant action, presumably affording low-level control over reaching and grasping for objects in peripersonal space (Romo, Hernandez, & Zainos, 2004; Wise, Di Pellegrino, & Boussaoud, 1992). Subsequently, neuroimaging experiments have drawn a link between simple visuomotor transformation and the premotor cortex (Jenkins, Jahanshahi, Jueptner, Passingham, & Brooks, 2000).

Posterior lateral PFC The function of the posterior lateral prefrontal cortex remains more controversial, but once again early lesion studies in monkeys forged the path to understanding its primary function, with the observation that removal of BA 8 impairs visuomotor selection that is contingent on a contextual cue (Petrides, 1985). Moreover, neurons found in posterior lateral portions of the macaque PFC seem to enact a triangular integration of stimulus, response, and task, precisely as if they were encoding how stimulus-response relations vary according to the currently active task or rule (Averbeck, Sohn, & Lee, 2006; Wallis, Anderson, & Miller, 2001). In humans, tasks that require decisions about "bivalent" stimuli, to which one of two possible responses is indicated by a concurrent (or immediately preceding) contextual cue—which are a feature of many experiments exploring the neural correlates of switching between tasks, or the neural basis for allocating attentional priority—invariably engage

the posterior dorsolateral prefrontal cortex, also known as the inferior frontal junction (Brass, Derrfuss, Forstmann, & von Cramon, 2005; Derrfuss, Brass, Neumann, & von Cramon, 2005). Posterior lateral PFC regions are also particularly sensitive to situations where a weak stimulus-response association must be favored over a stronger one, as is required for conflict resolution in the Stroop task (Kerns et al., 2004). All these effects can be conceived as instances of contextual control, where decision making demands the integration of neural signals linking stimulus and response with other information denoting the currently active task or rule (Bunge, 2004).

Midlateral PFC Third, in contradistinction to posterior lateral PFC regions, surgical removal of the more anterior lateral prefrontal zone (BA9/46) results in impaired performance on delayed "nonmatch to sample" tasks, in which information about an object or spatial configuration must be actively sustained across a delay period, to permit the selection of an alternate or nonmatching item (Petrides, 1991). Accordingly, neuroimaging studies tracking brain activity across the delay period on such tasks typically observe activations in midlateral portions of the prefrontal cortex, in BA9/46 (Curtis & D'Esposito, 2003; Wager & Smith, 2003). This region is probably the human homologue of the macaque periprincipal area in which tonically firing "memory" cells were first isolated by Fuster and his colleagues (Fuster & Alexander, 1971). There is thus considerable evidence underlining the role of the midlateral PFC in guarding task-relevant information active in a temporally extended fashion, thus allowing past task instructions to modulate action selection across a behavioral episode (episodic control), over and above more immediate contextual or sensorimotor signals.

The most direct evidence for the cascade model, however, comes from neuroimaging experiments that have measured brain activity on blocks of trials tailored to conform precisely to the demands of sensorimotor, contextual, and episodic control (Badre & D'Esposito, 2007; Koechlin et al., 2003). For example, Koechlin and colleagues acquired brain images while subjects performed (1) a simple choice among stimulus-response contingencies, for example, a right button press to a red square and a left button press to a green square (requiring sensorimotor control); or (2) either a vowel/consonant or an uppercase/lowercase discrimination on singly presented Roman letters, with the relevant task indicated by the color of the stimulus (requiring contextual control); or these tasks where (3) color-response (or color-task) assignments, rather than occurring consistently across experimental blocks (i.e., red = right, green = left), cycle from block to block (e.g., for some blocks, cyan = left, yellow = right; for others, cyan = right, yellow = left), requiring subjects to encode, maintain, and implement episodic information derived from the instruction cue at the debut of each block (requiring episodic control). Crossing sensorimotor, contextual, and episodic tasks in a factorially designed neuroimaging experiment has confirmed that midlateral prefrontal, posterior lateral prefrontal, and premotor cortices subserve episodic, contextual, and sensorimotor control, respectively (Badre & D'Esposito, 2007; Koechlin et al., 2003).

Patterns of anatomical connectivity observed in the lateral prefrontal cortex of the macaque lend further plausibility to hierarchical accounts of cognitive control, as each subregion of the PFC seems to share connectivity both with higher and lower regions in the chain of command, as well as with corresponding regions in posterior neocortex (Barbas, 2000). Premotor cortex and posterior lateral and midlateral PFC are thus ideally placed to make complementary contributions to top-down control of perception and action, by means of reciprocal interconnectivity with zones responsible for perceptually treating and encoding the objects prompting the decision process. Moreover, assaying functional connectivity during the implementation of episodic and contextual control has revealed that coupling from relevant higher to lower regions is increased by each of these control processes, consistent with a backward cascade of information through the processing hierarchy during decision making (Koechlin et al., 2003).

An emerging view thus characterizes the decision processes leading to action selection in the LFC as a hierarchical system. In the cascade model, successively anterior regions are not homuncular command centers that orchestrate behavior in lower modules, but rather an integrated network that collectively extends the temporal frame across which a decision is taken. This permits information from the more and more distant past to be brought to bear upon a decision, allowing the flexible selection of action across diverse and rapidly changing contexts and episodes.

The medial frontal cortex: Movement, monitoring, and motivation

Neural activity in the medial frontal cortex (MFC) frequently accompanies LFC activity during decision making. The MFC comprises the gray matter extending along the median walls of the frontal lobe, from the motor cortex to the orbitofrontal cortex. The anterior cingulate cortex (ACC), also including caudal and rostral sectors, follows the length of the medial frontal cortex at its inferior aspect abutting the white fiber bundles of the corpus callosum, making up about half its total volume. In contrast to the emerging consensus about the functional organization of the LFC, however, much less is known about how different regions within the MFC might contribute to decision making. Here, drawing together a spectrum of current theories of MFC function, we argue that the primary role of the MFC is *motivational control*, that is, a

calculation of the requisite investment for optimal control over action selection. Subsequently, we will raise the possibility that there may be a hierarchical implementation of motivational control along the rostrocaudal axis of the MFC.

Anatomical Definitions For the purposes of this review, we divide the MFC into three sectors (Amodio & Frith, 2006; Koski & Paus, 2000; Picard & Strick, 1996). (1) The dorsal MFC (dMFC) comprises caudal motor areas and the SMA; (2) the posterior rostral MFC (prMFC) includes the pre-SMA and the dorsal portion of the ACC lying anterior to $y = 10$; (3) the anterior rostral MFC (arMFC) extends more anteriorly to the border with the orbitofrontal cortex, at $z > 0$, including regions often termed the paragenual and subgenual cingulate, and ventromedial prefrontal cortex. See figure 70.1 for details. Our discussion focuses largely upon sectors 1 and 2, and in much of what follows we use the term MFC to refer to the dMFC/prMFC.

Theories of MFC Function Historically, the MFC has been considered a critical part of the motor system, and early theories focused on the various contributions of the SMA and ACC to action planning and execution (Dum & Strick, 1992) and to converting intentions into actions (Paus, 2001). Subsequent work framed the ACC as part of the "executive" attention system of the human brain, as early PET studies revealed its participation in domain-independent target-detection processes (Posner & Petersen, 1990) or the acquisition of novel complex information (Raichle et al., 1994), giving rise to the view that the MFC may participate in "attention to action" (Passingham, 1996). However, it was the stream of research that began with the observation that the dorsal MFC responds vigorously to errors (Carter et al., 1998; Falkenstein, Hohnsbein, & Hoormann, 1995; Niki & Watanabe, 1979; Taylor, Stern, & Gehring, 2007) and situations involving conflict among competing response options (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999) that forms the basis for most contemporary thinking about the function of the MFC. Current perspectives, which will be summarized in more detail later, emphasize an *evaluative* role for the MFC, arguing that its primary function is to monitor and adjust levels of control needed for efficient implementation of task and action selection in the lateral PFC (Botvinick, 2007; Holroyd & Coles, 2002; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004; Rushworth & Behrens, 2008). This evaluative role might be important for maximizing outcome across an instrumental learning schedule (Holroyd & Coles, 2002; Rushworth & Behrens, 2008) and/or assessing the costs associated with an action (Walton, Rudebeck, Bannerman, & Rushworth, 2007), casting the ACC as a critical part of the reward system of the human brain, with a role in representing the subjective value of actions (Rushworth, Behrens, Rudebeck, & Walton, 2007). These approaches assume that control is a flexible resource whose allocation can be increased to meet the needs of a particularly challenging task or to encourage optimal performance in a situation where the stakes (potential gains or losses) are raised. We call this process by which investment is adjusted to modulate decision making *motivational control*, and we suggest that it may be implemented by way of recurrent interactions between the LFC and MFC.

Adjusting Motivation to Meet Changing Task Demands Many environments are inherently unstable, and levels of control required to arbitrate among possible actions in a stable, familiar milieu are unlikely to suit efficient decision making in a novel or unpredictable situation (Behrens, Woolrich, Walton, & Rushworth, 2007). For example, even a seemingly tranquil, uneventful drive along the open motorway can abruptly require higher levels of control to be engaged, for example, when sudden heavy rain reduces visibility, the traffic bunches up, and you are diverted through an unfamiliar section of the carriageway. Neuroimaging studies are often designed to mimic this volatility, intermixing high-demand and low-demand trials in rapid succession (such as "congruent" and "congruent" in the Stroop task, or "switch" and "stay" trials in task-switching paradigms). Noting the striking sensitivity of medial prefrontal regions to the onset of conflict among competing stimulus-response assignments (Botvinick et al., 1999), as well as to negative performance outcomes, such as errors (Carter et al., 1998; Gehring, Gross, Coles, Meyer, & Donchin, 1993), many theorists have argued that the MFC keeps track of the likely demand associated with a forthcoming event or behavioral episode (Brown & Braver, 2005), perhaps signaling this information to the lateral PFC in order to dynamically adjust cognitive control according to current needs. Evidence in favor of this view comes from studies reporting neural signals in the MFC that precede and predict task-induced adjustments in performance. Among the best described of these is the speed-accuracy trade-off, whereby challenging or high-risk situations tend to be approached more cautiously, with slower but more accurate responses. For example, the slowing in reaction times that immediately follows an error trial ("posterror slowing") (Rabbitt, 1966) is predicted by the amplitude of the error-related negativity (ERN), a potential with a probable source in the MFC (Gehring et al., 1993; Yeung, Cohen, & Botvinich, 2004). Similarly, activity in the ACC predicts "conflict adaptation" (Kerns et al., 2004), that is, the reduced response-time cost on an incongruent trial following another incongruent trial (Gratton, Coles, & Donchin, 1992), as well as predicting LFC activity on the subsequent trial. Conflict adaptation may occur because

levels of control are increased once conflict has been detected, such that the second incongruent trial is tackled more effectively than the first (Egner, 2007).

Adjusting Motivation to Meet Changing Action-Outcome Associations In addition to these moment-to-moment fluctuations in the level of cognitive resources or effort required to tackle the local environment, so also the risk, stake, and value associated with actions are subject to change. Adjustments to the gains and losses associated with an action might be gradual, as with encroaching satiety during feeding, but they can also reverse abruptly; for example, the slot machines at an amusement park might yield nothing for several successive trials, but then suddenly offer a large payout. Accumulating evidence suggests that the MFC can step in to increase control on the basis of shifting reward contingencies, allowing motivational signals to guide decision making on the basis of likely reward and punishment (Holroyd & Coles, 2002; Rushworth, Walton, Kennerley, & Bannerman, 2004). Notably, large portions of the MFC respond to aversive stimuli, such as pain (Vogt, 2005), but primary (Amiez, Joseph, & Procyk, 2006) and secondary (McClure, Laibson, Loewenstein, & Cohen, 2004) rewards also evoke potent responses in the posterior MFC, with cells responsive to positive and negative feedback often intermingled in the same stretch of tissue (Matsumoto, Matsumoto, Abe, & Tanaka, 2007). However, just as for negative-valenced or conflict-based signals, this response to positive feedback is often contingent on its relevance for future behavior. For example, neurons in the MFC respond to an unexpected reduction in the level of reward associated with a given action—but only when the monkey takes this event as a cue to switch to another movement, in order to maximize overall outcome (Cohen & Ranganath, 2007; Shima & Tanji, 1998). The responses of both neurons that prefer positive feedback and those that prefer negative feedback are scaled by the disparity between the reward that is expected and that which is actually received—with greater signals elicited by more surprising events in both cases (Matsumoto et al.; Oliveira, McDonald, & Goodman, 2007). This evidence has given rise to an emerging consensus that the MFC monitors for the mismatch between expected and observed outcomes—whether positive or negative—adjusting motivational control in order to maximize the benefits and minimize the costs associated with a chosen course of action.

Functional fractionation of the MFC

Does the MFC contribute to motivational control monolithically, or are dissociable contributions made by different subsectors, such as the dMFC, prMFC, and arMFC? In this final section we propose a theory of functional organization of the MFC, inspired by the intuition that motivational control signals may be engaged on the basis of current, contextual information, such as the immediate presence of two stimuli associated with conflicting response tendencies ("contextual motivational control") or may be required to adjust decision making over a longer episode with particular demand or outcome characteristics ("episodic motivational control"). Based on a review of the extant literature and on new work from our own laboratory, we propose that episodic and contextual motivational control are implemented in rostral (i.e., dorsal ACC) and caudal portions (i.e., SMA) of the posterior MFC, respectively (figure 70.2).

Initial support for this view comes from extant studies describing the implementation of conflict and error processing in the MFC. Although some researchers have suggested that the MFC response to errors on tasks requiring high levels of control (such as the Stroop) might mask a primary sensitivity to the response competition that provoked the error (Carter, Botvinick, & Cohen, 1999), it is interesting to reflect that response conflict (once detected) must be resolved immediately in order for an optimal decision to be made, whereas errors or negative feedback invariably occur too late to apply to immediate behavior—rather, they are relevant to future events occurring within the same episode (Ullsperger & von Cramon, 2004). Indeed, it is intriguing (but rarely noted) that posterror slowing and conflict adaptation have opposing effects on response time, consistent with the view that errors and conflict may have dissociable neurocognitive consequences. One possibility, therefore, is that resolving immediate response conflict thus demands contextual motivational control, but errors may lead to an up-regulation of episodic motivational control.

This contention is supported by a growing literature reporting that response conflict is primarily the province of more posterior, dorsal MFC regions, such as the SMA (and/or more anterior counterpart, the pre-SMA) whereas errors tend to activate the ACC proper (Nachev, 2006; Rushworth et al., 2004; Ullsperger & von Cramon, 2004), even in studies where the overlap between conflict- and error-sensitive voxels is emphasized (Kerns et al., 2004). One influential review (Rushworth et al., 2004) notes that across a number of fMRI studies in which response conflict was elicited—including versions of the oddball, flanker, and go-no-go tasks—error responses were isolated in unambiguous cingulate territory, whereas conflict without error additionally activated the motor zones of the superior frontal gyrus. Similarly, evoked scalp negativities recorded over the MFC of a patient with a lesion confined to the ACC dissociated error-monitoring responses from conflict-detection responses (Swick & Turken, 2002). These findings suggest that the SMA and ACC may make dissociable contributions to the processing of conflict and errors, respectively, and support

the view that the ACC is activated when motivational control is required over an extended episode, whereas the dorsal MFC is responsible for energizing control processes in response to immediate contextual incentives.

Immediate and Remote Relevance of Demand Is this perspective at variance with the widely held view that it is the ACC, not the SMA, that is responsible for conflict detection? We find it useful to distinguish activations that predict trial-by-trial adjustments in the neurocognitive correlates of control on the *subsequent* trial and those that simply respond to a cognitive challenge on the *current* trial, such as voxels sensitive to incongruent events on a Stroop paradigm. Neural correlates of adjustment to future events, such as those observed during conflict adaptation, tend to occur in the ACC but not the SMA (Botvinick et al., 1999), whereas the simple demand for response competition to be addressed—such as is found on incongruent, relative to congruent, trials—activates a swath of medial frontal cortex extending well beyond the ACC into the pre-SMA and SMA (Barch et al., 2001; Nee, Wager, & Jonides, 2007). Other studies have specifically noted that the ACC is activated by motivational information that is relevant to forthcoming events, such as the probability that an error will be made on future trials (Brown & Braver, 2005). Moreover, in task-switching paradigms, where an entire experimental block is associated with increased likelihood of a switch, the ACC but not the SMA is engaged in a sustained fashion (Braver, Reynolds, & Donaldson, 2003; Dosenbach et al., 2006). These findings are wholly consistent with the view that the ACC tracks the need for motivational control across a future episode, in contrast to the SMA, which responds to conflict among current motor plans (Nachev, Wydell, O'Neill, Husain, & Kennard, 2007).

Immediate and Remote Relevance of Reward A substantial literature describing single-cell electrophysiology data in awake, behaving monkeys suggests that ACC neurons encode episodic aspects of reward—those whose relevance extends over the longer term. For example, the reward responsivity of ACC (but not SMA) neurons is dependent on whether the outcome triggers a change in strategy, such as a switch from one action to another (Shima & Tanji, 1998), and ACC neurons signal with responses of increasing potency the proximity of a reward obtained at the end of an extended behavioral sequence (Shidara & Richmond, 2002), and may encode an estimate of the value associated with an ongoing task (Amiez et al., 2006; Sallet et al., 2007). Perhaps the most striking evidence, however, comes from lesion studies in monkeys with comprehensive cingulate lesions, who exhibit relatively normal motor control but show deficits on a decision-making task that requires the reward history to be integrated across several trials to optimally select a response (Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006; Walton et al., 2007). The neural sequelae of reward are not so well characterized in human dorsal MFC, but it is intriguing that fMRI activity in the ACC tracks temporal discounting, the relative value of a reward as its arrival is delayed into the future (McClure, Ericson, Laibson, Loewenstein, & Cohen, 2007; McClure et al., 2004). Moreover, recent studies that have combined computational modeling with neuroimaging have noted the sensitivity of the ACC to parameters encoding variation in the reward rate associated with a response, tantamount to an estimate of the ongoing volatility of the environment (Behrens et al., 2007) or to the tendency for the organism to explore available options rather than exploit a reliable reinforcer (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). These results imply that the ACC is particularly active whenever the environment contains rich information about the reinforcement available in the current episode (i.e., a higher learning rate), a contention that is consistent with its extensive interconnectivity with other limbic structures involved in learning and memory (Vogt, Pandya, & Rosene, 1987).

Direct evidence for a hierarchy of motivational control, however, comes from an fMRI study in which contextual and episodic motivational control were varied as independent factors (Kouneiher, Charron, & Koechlin, in press). During performance of a challenging cognitive task, subjects were financially rewarded for each block of 12 trials completed with no errors (+1 euro), received no reward for blocks with a single error, and were punished incrementally for each additional error (−1 euro). However, specific trials (bonus trials) designated with a contextual signal (a frame around the stimulus) carried additional incentives, which we reasoned would lead to increases in contextual motivational control. Critically, however, the stake associated with these additional incentives was scaled on a block-by-block (episodic) basis: on high-episodic-motivation blocks performance in bonus trials led to additional gains/losses (±2 euros), whereas on low-episodic-motivation blocks, the additional gains/losses were negligible (±5 cents). The neuroimaging results supported the proposed dissociation between episodic and contextual motivational signals: the ACC responded in a sustained fashion to high > low episodic motivation, with a response that deviated reliably from zero even on non-bonus trials falling within a high-episodic-motivation block. By contrast, a region falling on the SMA/preSMA border phasically responded to only bonus trials in high-motivation blocks, consistent with its participation in the processing of immediate contextual incentives. This dissociation between caudal and rostral subportions of the posterior MFC awaits confirmation by further studies.

Anterior Rostral MFC: Outcomes Associated with Hypothetical States? A discussion of the functional

fractionation of the MFC would not be complete without consideration of the arMFC, which seems to constitute yet another functionally distinct subdivision of the MFC (Steele & Lawrie, 2004), notable for its participation in the cognitive processes underlying social behavior, such as evaluating the mental states of others (Amodio & Frith, 2006) or contemplating the self (Mitchell, Banaji, & Macrae, 2005). These findings are complemented by parallel studies implicating the arMFC in decisions based on emotional information (Bush, Luu, & Posner, 2000; Etkin, Egner, Peraza, Kandel, & Hirsch, 2006) and the regulation of emotional responding (Ochsner, Bunge, Gross, & Gabrieli, 2002), leading to the conjecture that control operates in neuroanatomically distinct (dorsal) "cognitive" and (rostral) "emotional" domains. However, in addition to their role in social/emotional processing, the arMFC and accompanying subgenual cingulate cortex are also sensitive to a perplexing variety of other cognitive and motivational factors, including autobiographical recollection and imagination (Gehring & Knight, 2000), and decision processes that require the estimation of the outcome associated with an ambiguous situation or "generative model" of the world. For example, lesions to the arMFC engender deficits on "reversal learning" tasks in which stimuli with a high or low probability of reward are switched unpredictably (Fellows & Farah, 2007), forcing subjects to formulate an internal model of which of two possible states is currently active (Daw et al., 2006; Hampton, Bossaerts, & O'Doherty, 2006; Koechlin, Danek, Burnod, & Grafman, 2002). Moreover, arMFC lesions impair economic decisions where short-term gain must be subsumed in favor of a response program that maximizes gain over the longer term (Fellows & Farah, 2005). We speculate that the hierarchy of motivational control in the MFC continues into anterior rostral regions, where an outcome value may be assigned to a hypothetical future or "model-based" state rather than a currently ongoing episode—including the assignment of likely outcome to the mental states of others (Amodio & Frith, 2006). This distinction would mirror the dissociation observed in the LFC between BA 9/46 and the frontopolar cortex (BA 10), the latter of which seems to form a "buffer" that is engaged whenever a currently active task set is placed in a pending state, allowing flexible scheduling among different tasks (Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999; Koechlin & Hyafil, 2007) and thereby circumventing the well-described capacity limits ("attentional bottleneck") on central processing (Pashler, 2000; Sigman & Dehaene, 2005). The most rostral cortical areas thus may subserve the cognitive and motivational processes that allow us to interleave several behavioral episodes associated with different decision rules, allowing us to simultaneously perform two tasks with minimal mutual interference and maximal common outcome.

Conclusions

Great strides have been taken in understanding how the frontal lobes contribute to decision making, but much remains to be discovered. A crucial question for future research is likely to be the nature of the interactions between medial and lateral prefrontal cortices during decision making. While many theories have argued that MFC computes the motivational significance of an event and transmits this information to the LFC for task implementation (Carter et al., 1999), this unidirectional flow of information between MFC and LFC has been questioned by a study demonstrating that the error-related negativity is altered in patients with LFC lesions (Gehring & Knight, 2000). Instead, we argue that MFC and LFC may engage in a reciprocal exchange of information, allowing the relative importance of sensorimotor, contextual, and episodic information to be weighted as a function of its likely motivational significance. This process would allow the frontal lobes to arbitrate among potentially competing past and present influences upon action selection, in the interest of optimally guiding decision making.

ACKNOWLEDGMENTS We would like to thank Tobias Egner, Alexandre Hyafil, and Frederique Kouneiher for helpful comments on the manuscript. This work was supported by a European Young Investigator award to E.K.

REFERENCES

Amiez, C., Joseph, J. P., & Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. *Cereb. Cortex*, *16*, 1040–1055.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nat. Rev. Neurosci.*, *7*, 268–277.

Averbeck, B. B., Sohn, J. W., & Lee, D. (2006). Activity in prefrontal cortex during dynamic selection of action sequences. *Nat. Neurosci.*, *9*, 276–282.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J. Cogn. Neurosci.*, *19*, 2082–2099.

Barbas, H. (2000). Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices. *Brain Res. Bull.*, *52*, 319–330.

Barch, D. M., Braver, T. S., Akbudak, E., Conturo, T., Ollinger, J., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of response modality and processing domain. *Cereb. Cortex*, *11*, 837–848.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.*, *10*, 1214–1221.

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*, 179–181.

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cogn. Affect. Behav. Neurosci.*, *7*, 356–366.

Brass, M., Derrfuss, J., Forstmann, B., & von Cramon, D. Y. (2005). The role of the inferior frontal junction area in cognitive control. *Trends Cogn. Sci.*, *9*, 314–316.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, *39*, 713–726.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*, 1118–1121.

Bunge, S. A. (2004). How we use rules to select actions: A review of evidence from cognitive neuroscience. *Cogn. Affect. Behav. Neurosci.*, *4*, 564–579.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.*, *4*, 215–222.

Carter, C. S., Botvinick, M. M., & Cohen, J. D. (1999). The contribution of the anterior cingulate cortex to executive processes in cognition. *Rev. Neurosci.*, *10*, 49–57.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–749.

Christoff, K., & Gabrieli, J. D. E. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology*, *28*, 168–186.

Cohen, M. X., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *J. Neurosci.*, *27*, 371–378.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.*, *7*, 415–423.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–879.

Derrfuss, J., Brass, M., Neumann, J., & von Cramon, D. Y. (2005). Involvement of the inferior frontal junction in cognitive control: Meta-analyses of switching and Stroop studies. *Hum. Brain Mapping*, *25*, 22–34.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, *18*, 193–222.

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., Burgund, E. D., Grimes, A. L., Schlaggar, B. L., & Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron*, *50*, 799–812.

Dum, R. P., & Strick, P. L. (1992). Medial wall motor areas and skeletomotor control. *Curr. Opin. Neurobiol.*, *2*, 836–839.

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cogn. Affect. Behav. Neurosci.*, *7*, 380–390.

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., & Hirsch, J. (2006). Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron*, *51*, 871–882.

Falkenstein, M., Hohnsbein, J., & Hoormann, J. (1995). Event-related potential correlates of errors in reaction tasks. *Electroencephalogr. Clin. Neurophysiol. Suppl.*, *44*, 287–296.

Fellows, L. K., & Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cereb. Cortex*, *15*, 58–63.

Fellows, L. K., & Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cereb. Cortex*, *17*, 2669–2674.

Fuster, J. M. (2000). Executive frontal functions. *Exp. Brain Res.*, *133*, 66–70.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, *173*, 652–654.

Gehring, W. J., Gross, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. A. (1993). Neural systems for error detection and compensation. *Psychol. Sci.*, *11*, 1–6.

Gehring, W. J., & Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nat. Neurosci.*, *3*, 516–520.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *J. Exp. Psychol. Gen.*, *121*, 480–506.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.*, *26*, 8360–8367.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.*, *109*, 679–709.

Jenkins, I. H., Jahanshahi, M., Jueptner, M., Passingham, R. E., & Brooks, D. J. (2000). Self-initiated versus externally triggered movements. II. The effect of movement predictability on regional cerebral blood flow. *Brain*, *123*(Pt. 6), 1216–1228.

Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J. & Rushworth, M. F. (2006). Optimal decision-making and the anterior cingulate cortex. *Nat. Neurosci.*, *9*, 940–947.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, *399*, 148–151.

Koechlin, E., Danek, A., Burnod, Y., & Grafman, J. (2002). Medial prefrontal and subcortical mechanisms underlying the acquisition of motor and cognitive action sequences in humans. *Neuron*, *35*, 371–381.

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science*, *318*, 594–598.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*, 1181–1185.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.*, *11*, 229–235.

Koski, L., & Paus, T. (2000). Functional connectivity of the anterior cingulate cortex within the human frontal lobe: A brain-mapping meta-analysis. *Exp. Brain Res.*, *133*, 55–65.

Kouneiher, F., Charron, S., & Koechlin, E. (in press). Motivation and cognitive control in the human prefrontal cortex. *Nat. Neurosci.*

Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.*, *10*, 647–656.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *J. Neurosci.*, *27*, 5796–5804.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503–507.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.*, *24*, 167–202.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *J. Cogn. Neurosci.*, *17,* 1306–1315.

Nachev, P. (2006). Cognition and medial frontal cortex in health and disease. *Curr. Opin. Neurol.*, *19,* 586–592.

Nachev, P., Wydell, H., O'Neill, K., Husain, M., & Kennard, C. (2007). The role of the pre-supplementary motor area in the control of action. *Neuroimage*, *36*(Suppl. 2), T155–163.

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cogn. Affect. Behav. Neurosci.*, *7,* 1–17.

Niki, H., & Watanabe, M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Res*, *171,* 213–224.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. (2002). Rethinking feelings: An FMRI study of the cognitive regulation of emotion. *J. Cogn. Neurosci.*, *14,* 1215–1229.

Oliveira, F. T., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action-outcome associations. *J. Cogn. Neurosci.*, *19,* 1994–2004.

Pashler, H. (2000). Task-switching and multitask performance. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of mental processes* (pp. 277–301). Cambridge, MA, MIT Press.

Passingham, R. E. (1988). Premotor cortex and preparation for movement. *Exp. Brain Res.*, *70,* 590–596.

Passingham, R. E. (1996). Attention to action. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *351,* 1473–1479.

Paus, T. (2001). Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nat. Rev. Neurosci.*, *2,* 417–424.

Petrides, M. (1985). Deficits in non-spatial conditional associative learning after periarcuate lesions in the monkey. *Behav. Brain Res.*, *16,* 95–101.

Petrides, M. (1991). Monitoring of selections of visual stimuli and the primate frontal cortex. *Proc. Biol. Sci.*, *246,* 293–298.

Petrides, M. (2005). Lateral prefrontal cortex: Architectonic and functional organization. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *360,* 781–795.

Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: A review of their location and functional activation. *Cereb. Cortex*, *6,* 342–353.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.*, *13,* 25–42.

Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *J. Exp. Psychol.*, *71,* 264–272.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. M., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb. Cortex*, *4,* 8–26.

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.*, *5,* 184–194.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306,* 443–447.

Romo, R., Hernandez, A., & Zainos, A. (2004). Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron*, *41,* 165–173.

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.*, *11,* 389–397.

Rushworth, M. F., Behrens, T. E., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behavior. *Trends Cogn. Sci.*, *11,* 168–176.

Rushworth, M. F., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends Cogn. Sci.*, *8,* 410–417.

Sakai, K., & Passingham, R. E. (2003). Prefrontal interactions reflect future task operations. *Nat. Neurosci.*, *6,* 75–81.

Sallet, J., Quilodran, R., Rothe, M., Vezoli, J., Joseph, J. P., & Procyk, E. (2007). Expectations, gains, and losses in the anterior cingulate cortex. *Cogn. Affect. Behav. Neurosci.*, *7,* 327–336.

Shallice, T. (1988). *From neuropsychology to mental structure.* New York: Cambridge University Press.

Shidara, M., & Richmond, B. J. (2002). Anterior cingulate: Single neuronal signals related to degree of reward expectancy. *Science*, *296,* 1709–1711.

Shima, K., & Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. *Science*, *282,* 1335–1338.

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biol.*, *3,* e37.

Steele, J. D., & Lawrie, S. M. (2004). Segregation of cognitive and emotional function in the prefrontal cortex: A stereotactic meta-analysis. *NeuroImage*, *21,* 868–875.

Swick, D., & Turken, A. U. (2002). Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex. *Proc. Natl. Acad. Sci. USA*, *99,* 16354–16359.

Taylor, S. F., Stern, E. R., & Gehring, W. J. (2007). Neural systems for error monitoring: Recent findings and theoretical perspectives. *Neuroscientist*, *13,* 160–172.

Ullsperger, M., & von Cramon, D. Y. (2004). Neuroimaging of performance monitoring: Error detection and beyond. *Cortex*, *40,* 593–604.

Vogt, B. A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nat. Rev. Neurosci.*, *6,* 533–544.

Vogt, B. A., Pandya, D. N., & Rosene, D. L. (1987). Cingulate cortex of the rhesus monkey. I. Cytoarchitecture and thalamic afferents. *J. Comp. Neurol.*, *262,* 256–270.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cogn. Affect. Behav. Neurosci.*, *3,* 255–274.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411,* 953–956.

Walton, M. E., Rudebeck, P. H., Bannerman, D. M., & Rushworth, M. F. (2007). Calculating the cost of acting in frontal cortex. *Ann. NY Acad. Sci.*, *1104,* 340–356.

Wise, S. P., Di Pellegrino, G., & Boussaoud, D. (1992). Primate premotor cortex: Dissociation of visuomotor from sensory signals. *J. Neurophysiol.*, *68,* 969–972.

Yeung, N., Cohen, J. D., & Botvinick, M. M. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychol. Rev.*, *111,* 931–959.

71 Circuits in Mind: The Neural Foundations for Object Concepts

ALEX MARTIN

ABSTRACT Functional neuroimaging studies have provided convincing evidence to support three main conclusions about the neural circuitry that underpins our understanding of objects in the world. First, our conceptual system contains property-based neural circuits *grounded* in the systems that support perceiving, acting, and feeling. Second, our conceptual system prominently includes relatively distinct neural circuits for processing and storing *domain-specific* information. Third, these circuits reflect the *interpretation* or meaning assigned to an object, not its physical features. Outstanding questions and problems with an embodied, domain-specific view of conceptual representation, as well as the role of the anterior regions of the temporal lobes in conceptual processing and semantic memory, are discussed.

Every day we encounter new exemplars of objects that we have never seen before. Yet we identify each as belonging to a particular category—as a chair, a dog, a tree—instantly and effortlessly. In fact, it has been shown that as soon as we *see* it we *know* what it is (Grill-Spector & Kanwisher, 2005). This mundane phenomenon underscores the fact that object recognition must be—in part—an act of memory. Perceiving, as William James recognized 120 years ago, is largely dependent on stuff that "comes out of our own heads." Indeed, for James, this idea was important enough to be considered "the general law of perception" (W. James, 1890). This chapter will focus largely on what we have learned about the stuff in our heads that allows us not only to perceive, but also to imagine and think about objects in the world.

What is an object concept?

For our purposes, *object concept* will be used to refer to the representation (i.e., the information stored in memory) of an object category (a class of objects in the external world) (Murphy, 2002). In this view, concepts are in our heads, categories are in the world. This distinction in no way undermines the fact that any object category (hammers, dogs) can be *categorized* in a nearly infinite variety of ways. For example, both hammers and dogs belong to the category of things that are smaller than a house. The neural basis for creating flexible ad hoc categories (Barsalou, 1989) will not be discussed here other than to note that the available neurophysiological evidence suggests that this ability rests heavily on activity in the prefrontal cortex, in interaction with the temporal lobes (see Miller, Nieder, Freedman, & Wallis, 2003, for review). Here I focus on the neural underpinnings for basic-level categories as defined in the following paragraphs.

The primary function of concepts is to allow us to quickly draw inferences about an object's properties. That is, identifying an object as, for example, a "hammer" means that we know that this is an object that is used to pound nails, so that we do not have to rediscover this property each time the object is encountered (see Murphy, 2002, for an extensive review of cognitive studies of concepts). In this sense, object perception involves not only making contact with stored information about the features present in the stimulus (e.g., what hammers typically look like), but also inferred information about other features or properties (e.g., those related to its function).

A major feature of object concepts is that they are hierarchically organized, with the broadest knowledge represented at the superordinate level, more specific knowledge at an intermediary level commonly referred to as the basic level, and the most specific information at the subordinate level. For example, "dog" is a basic-level category that belongs to the superordinate categories "animal" and "living things," and has subordinate categories such as "poodle" and "collie." As established by Eleanor Rosch and colleagues in the 1970s, the basic level has a privileged status (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Rosch, 1978). It is the level used nearly exclusively to name objects (e.g., "dog" rather than "poodle"). It is also the level at which we are fastest to verify category membership (i.e., we are faster to verify that a picture is a "dog" than an "animal" or a "poodle"). It is also the level at which subordinate category members share

ALEX MARTIN Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, Maryland

the most properties (e.g., collies and poodles have similar shapes and patterns of movement). Finally, the basic level is the easiest level at which to form a mental image (you can easily form an image of an elephant but not of an "animal"). This hierarchical organization has played a prominent role in the neuropsychology and computational modeling of semantic memory (e.g., McClelland & Rogers, 2003; and see chapter 72 in this volume by McClelland, Rogers, Patterson, Dilkina, & Lambon Ralph). Nevertheless, the great majority of neuropsychological and neuroimaging studies have concentrated on understanding how basic-level concepts are represented in the brain.

Neural foundations for conceptual representations

Before describing what we know about the circuitry underpinning the representation of basic object concepts, it is important to draw a distinction between explicit and implicit levels of knowledge representation and expression. There is no need for any organism to acquire information unless that information can be expressed. Organisms learn, and the evidence for that learning is demonstrated by a change in behavior. What is represented (stored) in the brain is information. What is expressed is knowledge. How this knowledge is expressed is of fundamental importance for understanding how information is represented. For humans, a primary, and arguably *the* primary mode of expression, is the language system. Questions designed to probe knowledge about a specific entity are posed orally or in written form, and subjects respond verbally. Occasionally, a manual response may be required (e.g., show me how you would use a hammer) either by actually manipulating the object or by pantomime. However, regardless of whether the response is verbal or manual, knowledge is expressed explicitly. This explicit knowledge is typically referred to as associative knowledge or encyclopedic knowledge, and it is this level that is typically probed in both normal and brain-damaged individuals. Associative or encyclopedic knowledge has three main characteristics. First, as noted earlier, retrieval is explicit. Second, there is no intrinsic limitation on the amount of information that can be stored and retrieved. For a specific category of objects, (e.g., *dogs*), we may know lots of things. We know they are living things, have four legs, are smaller than a car, like to take walks, like to play fetch, and so on, and so on. Moreover, it does not matter whether the information is true. If you believe that dogs can fly, then that information is part of your semantic knowledge about dogs and is represented somewhere in your brain. Finally, this level of knowledge is idiosyncratic. Some people know lots about dogs, whereas others know very little.

This explicitly expressed knowledge about objects can be contrasted with a different level of object concept representation referred to as core properties or "semantic primitives" (Martin, 1998). In contrast to encyclopedic knowledge, semantic primitives are accessed implicitly and automatically in the service of comprehension, are highly constrained in number, and are universal. This level of representation allows us to quickly and efficiently identify objects and understand words, and forms the foundation for our vast stores of encyclopedic knowledge about objects.

While the model to be described here does not address the organization of encyclopedic knowledge, it makes strong claims about the organization of semantic primitives with regard to both their representational content and organization in the brain. For example, the semantic primitives associated with common tools include stored representations of what they look like, how they move when used, and how they are manipulated. They are stored within the same neural systems active when we learned about those properties. Specifically, they are stored within visual processing systems for perceiving object form and object motion, as well as action systems responsible for visuomotor transformations and for grasping and manipulating objects. These primitives are assumed to underpin object meaning in perception—regardless of the stimulus modality (visual, auditory, tactile) or format (pictures, words)—and in thought and imagination.

The distinction between an implicit level and an explicit level of representation underscores the fact that the embodied view of conceptual representation to be discussed in this chapter is not meant to provide an exhaustive description of a concept. It is undoubtedly true that a great deal of what we know about any concept is mediated by, and stored in, the language system. As will be described, some of this information is directly grounded in perceiving, acting, and feeling (e.g., verbal information about sensory- and motor-system-based properties). Other types of information may be truly abstract (nonembodied) and verbally mediated only (although see Barsalou, 1999, for a different view of the relationship between abstract concepts and perceptual systems).

Object concepts are grounded in the neural systems that support perceiving, acting, and feeling

Embodied cognition, including the notion that object concepts are grounded in perception and action systems, has become an increasingly popular view in modern cognitive science (e.g., Barsalou, 1999, 2008; Lakoff & Johnson, 1999; Wilson, 2002; Zwaan & Taylor, 2006). Although it is new to cognitive science, this idea has, in fact, a long history in behavioral neurology. For example, in an article published in the first volume of the journal *Brain*, the neurologist W. H. Broadbent wrote, "The formation of an idea of any external object is the combination of the evidence respecting it received through all the senses" (Broadbent, 1878). This claim was echoed a number of years later by a young

Sigmund Freud in his classic monograph *On Aphasia*: "The idea of the object is a complex of associations composed of the most varied visual, auditory, tactile, and kinesthetic and other impressions" (Freud, 1891) (see figure 71.1). Of course, the idea of an object must include information obtained through the senses. Where else would the information come from? What made these claims nontrivial, however, was that for both authors the information they spoke of was located or stored in the sensory processing systems themselves. In their view, our concepts were not abstract, verbal information stored in a place (association cortex?), but rather concepts were directly grounded in our sensory systems (see also Lissauer, 1890/1988). This very modern view of embodiment stands in marked contrast to the view that dominated cognitive psychology since the downfall of behaviorism in the 1950s whereby concepts were considered to be abstract, propositional, and amodal (e.g., Anderson, 1983; for discussion see Barsalou, 1999). Strictly amodal formulations now have largely disappeared, largely because of neuropsychological and especially neuroimaging evidence. Thus, as recently stated by a prominent group of neuropsychologists specializing in the study of semantics, "Essentially all current theoretical positions about semantic memory share the view that much of the content of our semantic memory relates to perception and action, and is represented in brain regions that overlap with, or possibly even correspond to, the regions that are responsible for perceiving and acting" (Patterson, Nestor, & Rogers, 2007).

Most of the direct evidence to support this type of embodiment claim comes from neuroimaging studies. In one of the earliest attempts to explore this issue, we used positron emission tomography (PET) to measure brain changes when subjects verbally generate different types of object-associated properties. Subjects provided words denoting object-associated colors in one condition (e.g., "yellow" in response to an achromatic picture of a pencil), and the names of associated actions in another condition ("write" in response to that same object). In line with an embodied view, direct comparison of these conditions showed that generating color associates activated regions in the ventral temporal cortex, downstream from regions known to respond to low-level visual processing of object form and form-related properties like color, whereas verb generation produced activity in the lateral part of the temporal lobe just anterior to, and thus assumed to be downstream from, the region responsible for low-level visual motion processing (other regions were also selectively active, especially during verb generation; for details see Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995). The findings and conclusions were strengthened by the fact that the same results were found regardless of whether the stimuli were object pictures or their written names (Martin et al., 1995).

Nevertheless, there was a bit of hand waving here. The brain regions engaged during color perception and motion perception were not mapped, so the claim of embodiment—the correspondence between knowing and

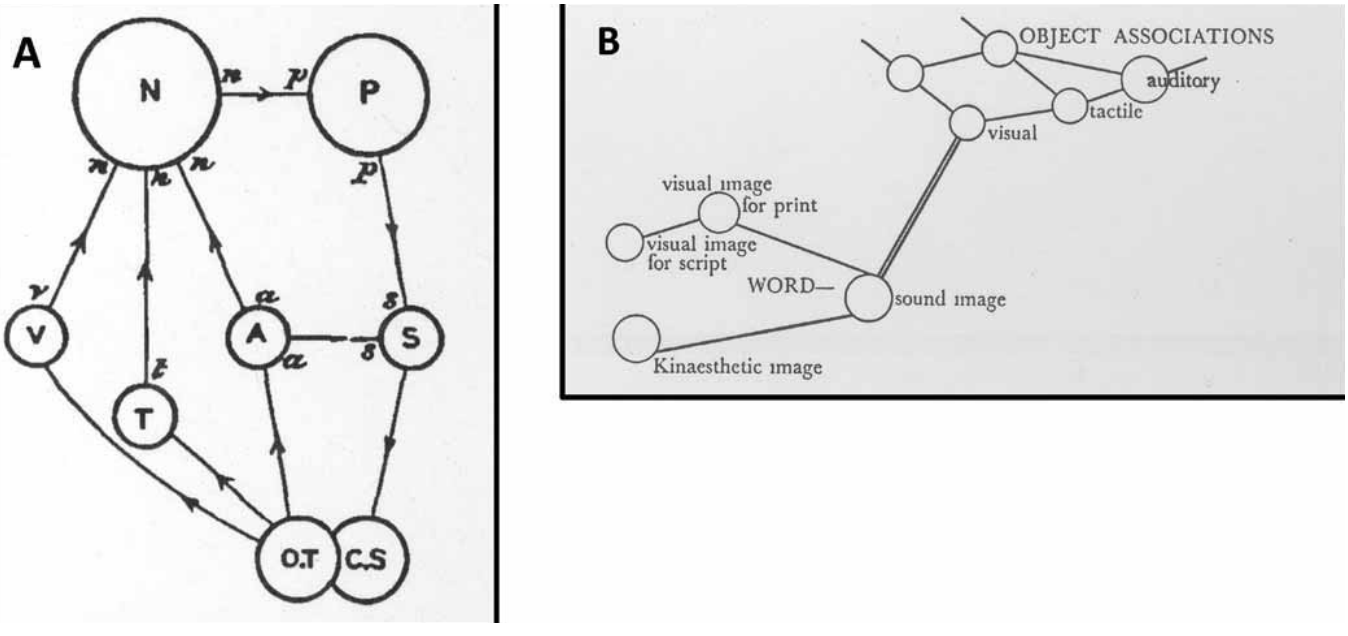


FIGURE 71.1 (*A*) Example of an embodied view of conceptual representation as depicted by W. H. Broadbent in 1878. *N* refers to the "Idea Centre" or "Naming Centre"; *V*, visual; *A*, auditory; *T*, tactile. *P* refers to "the propositional centre in which the phase was formed" (Broadbent, 1878). (*B*) Freud's diagram. He referred to this as "Psychological schema of the word concept" (Freud, 1891).

perceiving—was based on the presumed close spatial relationship between the activations elicited by the property-production tasks and the previously reported locations of the activity associated with color and motion processing. The embodied cognition view requires that the brain regions engaged when retrieving information about a sensory-based property like color overlap with the regions engaged when perceiving that property. In these initial studies overlap could not be determined. Moreover, subsequent attempts to directly evaluate this possibility failed to provide support. Rather, those data supported the initial conclusion that, although there was a close correspondence between the neural systems supporting perceiving and knowing (based on the location of their respective activations), they did not directly overlap (Chao & Martin, 1999). Consistent with previous reports (e.g., Zeki et al., 1991), viewing colors activated the lingual gyrus in occipital cortex, whereas retrieving information about color activated a more anterior region located in the fusiform gyrus in the posterior temporal lobes (Chao & Martin, 1999).

The finding that the neural substrates for perceiving and knowing were close but not overlapping could be used to undermine claims of embodiment (e.g., Mahon & Caramazza, 2008). After all, "close" is a relative term, and there is certainly no guarantee that there is any processing relationship between regions located a centimeter or more apart on the cortical surface (Chao & Martin, 1999). More recent evidence, however, has resolved this apparent problem by showing a direct overlap in the neural bases of perceiving and knowing. This result was accomplished by using a more demanding perceptual task than the passive viewing tasks previously employed to map sensory processing systems.

In a study on color perception, Beauchamp, Haxby, Jennings, and DeYoe (1999) reported activation in the lingual gyrus of the occipital cortex using a passive viewing task. This finding replicated previous neuroimaging studies, as noted previously. However, when the task was made more demanding by requiring subjects to judge subtle differences in hue, activity associated with perceiving color now extended downstream from the occipital cortex into the fusiform gyrus on the ventral surface of the temporal lobe. Thus the full extent of the color-processing system was revealed when the task was made more demanding, even though the same stimuli were used in both the passive-viewing and attention-demanding contexts (Beauchamp et al., 1999). Simmons, Ramjee, McRae, Martin, and Barsalou (2007) took advantage of this procedure to once again address the question of whether there was neural overlap between the systems underpinning perceiving and knowing about a specific object property. Using the attention-demanding hue-judgment task to evaluate color perception, and a verbal property-verification task to assess property knowledge, Simmons and colleagues found that retrieving information about object color—but not object motion—did, in fact, activate the same region in the fusiform gyrus active when color is perceived (Simmons et al., 2007) (figure 71.2). Thus, in support of the embodied concept view, these data provide strong evidence that information about a particular object property, like its typical color, is stored in the same neural system active when that property is perceived.[1]

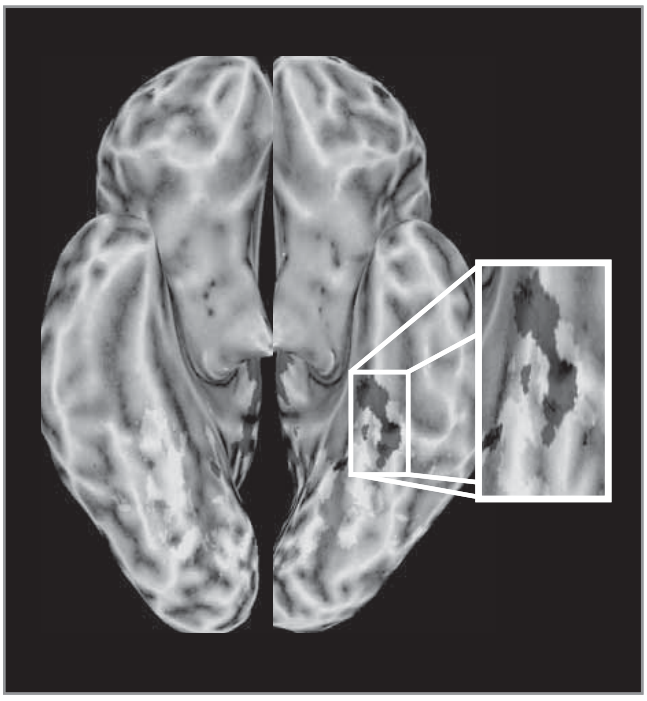


Figure 71.2 Overlap between the neural circuitry for perceiving and knowing about color. Shown is an inflated map of the ventral surface of the brain. Regions shown in yellow were more active when subjects performed a difficult color-perception task, relative to performing that same task with gray-scale stimuli. Regions in blue were more active when answering written questions about object color, relative to answering questions about object motor and motion properties. Red shows region of overlap in the left fusiform gyrus for the color-perception and color-knowledge tasks. (Adapted from Simmons, Ramjee, McRae, Martin, & Barsalou, 2007.) (See color plate 84.)

There are now many examples to support this claim (for extensive recent reviews and discussion see Barsalou, 2008; Gallese & Lakoff, 2005; Martin, 2007; Thompson-Schill, Kan, & Oliver, 2006). Examples include studies showing that retrieving information about different object-associated sensory properties (how they look, sound, feel, and taste) activated regions associated with sensory processing in each of these modalities (Goldberg, Perfetti, & Schneider, 2006), that making semantic judgments about words referring to body movements activated a region involved in perceiving biological motion (posterior region of the STS; Noppeney, Josephs, Kiebel, Friston, & Price, 2005), that reading emotionally charged words activated regions involved in perceiving emotions (amygdala, Kensinger &

Corkin, 2004), and that viewing pictures of appetizing foods activated gustatory processing and taste-specific reward areas (insula and orbitofrontal cortex, Simmons, Martin, & Barsalou, 2005). (It should be noted, however, that the majority of these studies did not independently localize the target sensory processing system, but rather relied on previously published localization studies.)

Similar findings have also been reported for the motor system. In perhaps the best known example, Pulvermuller and colleagues reported that simply reading words referring to actions performed with a particular body part (e.g., *lick*, *kick*, *pick*) activated corresponding regions in premotor and motor cortex (e.g., face, foot, and hand representations, respectively, as directly mapped by a movement study; Hauk, Johnsrude, & Pulvermuller, 2004). However, the correspondence between the primary motor representation for a specific body part (leg) and a concept associated with moving that same body part (*kick*) may be problematic, as will be discussed later (and see Mahon & Caramazza 2005, 2008, for insightful and penetrating critiques of the problems with some strong versions of the embodied viewpoint).

These findings underscore two important and related points. The first point concerns the need to distinguish between the neural bases for *sensation* and *perception* (Mesulam, 1998). As supported by the findings described previously, color *sensation* (color detection), as assessed by passive viewing, seems to be mediated by regions of occipital cortex located early in the visual processing stream, whereas color *perception* seems to require more extensive neural activity extending downstream into the fusiform gyrus. This distinction, in turn, fits nicely with the clinical literature that has documented a double dissociation between acquired color blindness (achromatopsia)—most commonly caused by a lesion of the lingual gyrus in the occipital lobes (Zeki, 1990)—and color agnosia—most commonly associated with lesions of posterior, ventral temporal cortex (Shuren, Brott, Schefft, & Houston, 1996). In this view, the posterior region in the lingual gyrus would be necessary for color sensation—detecting color and delivering this information to the rest of the processing system—whereas full perception of color—the experience of color bound to objects in the world—would require participation of more anterior regions. This anterior site may also provide the neural substrate for acquiring new object-color associations and representing those associations in memory.

The second important point concerns the fact that the overlap between perceiving and knowing is limited to only part, and in this case the most anterior part, of the sensory processing system. The claim then is not that conceptual information is stored throughout the entire sensory or motor processing system. Rather, the claim is that there is overlap between portions of these systems. This is an important point. Strong versions of embodied concept representation that can be construed as maintaining that concepts are grounded in the early stages of perception (V1) or motor processing system (M1) are vulnerable to a charge of reductio ad absurdum. (For example, with regard to the representation of action concepts, why not include the spinal cord? Why not include the muscles? See Mahon & Caramazza, 2005.) Formulations of embodiment that include primary sensory and motor cortices as part of the conceptual system must also account for why we do not move when we read the word "kick." They also need to explain how we are able to tell the difference between our visual perception of objects in the world and our visual imagery. In the current formulation, the overlap between the systems underpinning perceiving, acting, and knowing is limited. The overlap is partial, not complete. Information about a specific object property is stored in the anterior aspects of systems that are also active when objects are perceived and manipulated. This feature accounts for clinical dissociations and guards against a reductio ad absurdum argument while maintaining an embodied view. By so doing, however, the format and nature of the stored representations remain an open question. I will return to this issue at the end of the chapter.

Property information is organized in domain-specific neural circuitry

The most important point to come out of the studies that I have discussed, as well as from a wealth of neuropsychological investigations dating back well over 100 years, is that conceptual knowledge is not stored in a single location. The information that underpins our ability to know about our world is distributed throughout the brain. There is no single semantic memory store. Moreover, much of the available evidence suggests that this information is organized into relatively distinct, but broadly defined, domain-specific systems (e.g., Caramazza & Shelton, 1998). For our present purposes, a domain-specific system will refer to an information processing and storage system defined by the *type* or category of information it processes. These systems are composed of discrete cortical regions wired together to form relatively stable neural circuits. It is further assumed that the connections between the nodes or regions in these circuits are, in part, genetically predetermined. Different brain regions are predisposed to form connections with one another.

Motivated by the clinical literature on category-specific knowledge deficits, perhaps the most investigated domain-specific neural systems have been those concerned with representing animate entities, defined as living things that move on their own (people and other animals; Chao, Haxby, & Martin, 1999), and manipulable, manmade objects like common "tools," defined as objects with a systematic relationship between their visual form and function/manipulation (Mahon et al., 2007). The neural substrate for

representing animate things includes two regions of posterior temporal cortex; one of these lies on the ventral surface and is located in the more lateral portion of the fusiform gyrus (including, but not limited to, the fusiform face area, FFA, and its adjacent region for body representation; Schwarzlose, Baker, & Kanwisher, 2005); the other is on the lateral surface located in the posterior region of the superior temporal sulcus (pSTS). There is a substantial body of literature linking these regions to the representation of biological form (lateral portion of the fusiform gyrus) and biological motion (pSTS) (for reviews see Adolphs, 2001; Bookheimer, 2002; Haxby, Hoffman, & Gobbini, 2000; Martin, 2001, 2007; Martin & Chao, 2001; Thompson-Schill, 2003). This circuitry also includes the amygdala. The amygdala is a highly differentiated structure and plays multiple roles in emotion processing and behavior, including being critical for acquiring, storing, and expressing conditioned fear responses (Phelps & LeDoux, 2005; Phelps, 2006). In addition, there is growing evidence that the amygdala is predisposed to respond automatically to animate things. This is especially so for faces expressing fear, but the amygdala also responds more to neutrally posed faces relative to other objects (e.g., Pessoa, McKenna, Gutierrez, & Ungerleider, 2002), suggesting a predisposition for certain categories of objects over others (Ohman & Mineka, 2001). Indeed, recent studies from our laboratory suggest that the amygdala responds more to animate entities (faces and animals) than to other objects. Moreover, this response is especially strong for animate objects rated as being highly threatening and arousing (i.e., faces with expressions of fear, spiders, snakes), even when compared to equally threatening and arousing inanimate things (e.g., weapons, dental drills) (Yang, Bellgowan, & Martin, 2008). These data and others (e.g., Vuilleumier, Armony, Driver, & Dolan, 2003; G. Williams, Nestor, & Hodges, 2005) provide support for the inclusion of the amygdala in the circuitry for animate entities, both for assessing affective valence and arousal value, and for the fast, early detection of stimuli that have, from an evolutionary standpoint, posed the greatest threat—animals and other people.

The three regions listed—the lateral portion of the fusiform, pSTS, and the amygdala—respond strongly to both people and animals relative to other object categories. The available evidence further suggests that these regions code for different properties of animate things such as form, motion, and affective valence, respectively. Nevertheless, it should go without saying that the substrate for representing a property like visual form must distinguish between people and animals. All objects must have a distinct neural substrate, or how else would we distinguish among them? This distinction is clear in the clinical literature (Caramazza & Shelton, 1998). Thus, although a number of prosopagnosia patients also have difficulty identifying animals—for example, one of the most carefully studied prosopagnosic patients also was unable to identify animals from their shapes (patient LH; Etcoff, Freeman, & Cave, 1991)—convincing cases of pure prosopagnosia have been well documented (e.g., Riddoch, Johnston, Bracewell, Boutsen, & Humphreys, 2008). Thus these regions should be seen as part of the circuitry underpinning perceiving and knowing about animate entities, broadly defined, but with finer distinctions made between the representation of conspecifics and heterospecifics. Although how this distinction is represented in this circuitry has not been well defined, two hints are available. One hint comes from the neuroimaging literature that suggests that faces are more focally represented (Chao, Haxby, & Martin, 1999). This is not an unreasonable expectation given that different faces are highly homogeneous in shape and movement relative to animals, and they denote a single basic-level category, whereas animals are composed of stimuli with large variation in shape, and consist of multiple basic-level categories each with a unique name. The other hint comes from the clinical literature suggesting that they may be hemispheric difference, with a right-sided bias for lesions yielding face-processing deficits (Riddoch et al.) and a left-sided bias for lesions resulting in knowledge deficits for animals (Capitani, Laiacona, Mahon, & Caramazza, 2003).

In addition to the posterior, lateral region of the fusiform gyrus, pSTS, and the amygdala, other likely nodes in the animacy circuit include the medial portions of anterior and posterior cortex (ventral prefrontal and posterior cingulate/precuneus cortices, e.g., Mitchell, Macrae, & Banaji, 2006; Mitchell, 2008) and temporal polar cortices (Olson, Ploaker, & Ezzyat, 2007). Each of these nodes, along with a region located in posterior lateral cortex at the junction of the temporal and parietal lobes (Saxe, 2006), has been linked to rather abstract, higher-order aspects of social cognition, including the ability to make inferences about the mental state of others (with finer dissociations observed as well—for example, between different regions of medial prefrontal cortex when thinking about the mental states of similar versus dissimilar others; Mitchell et al.). Although some evidence exists that suggests that these regions may be involved in knowing about animacy in general (e.g., medial prefrontal cortex was found to be active when making judgments about mental states regardless of whether the target was another person or a dog; Mitchell, Banaji, & Macrae, 2005), most evidence suggests that these regions may be particularly important for thinking about conspecifics.

Each of the regions or nodes of this circuit has a specific function, and a major goal of cognitive neuroscience is to specify the functional properties of these regions in the service of social cognition. Clearly this issue is far from settled, and debate about the functional characteristics of each node is likely to continue for some time. Nevertheless, the critical point to be stressed here is that regardless of their function, each of the regions discussed so far is engaged

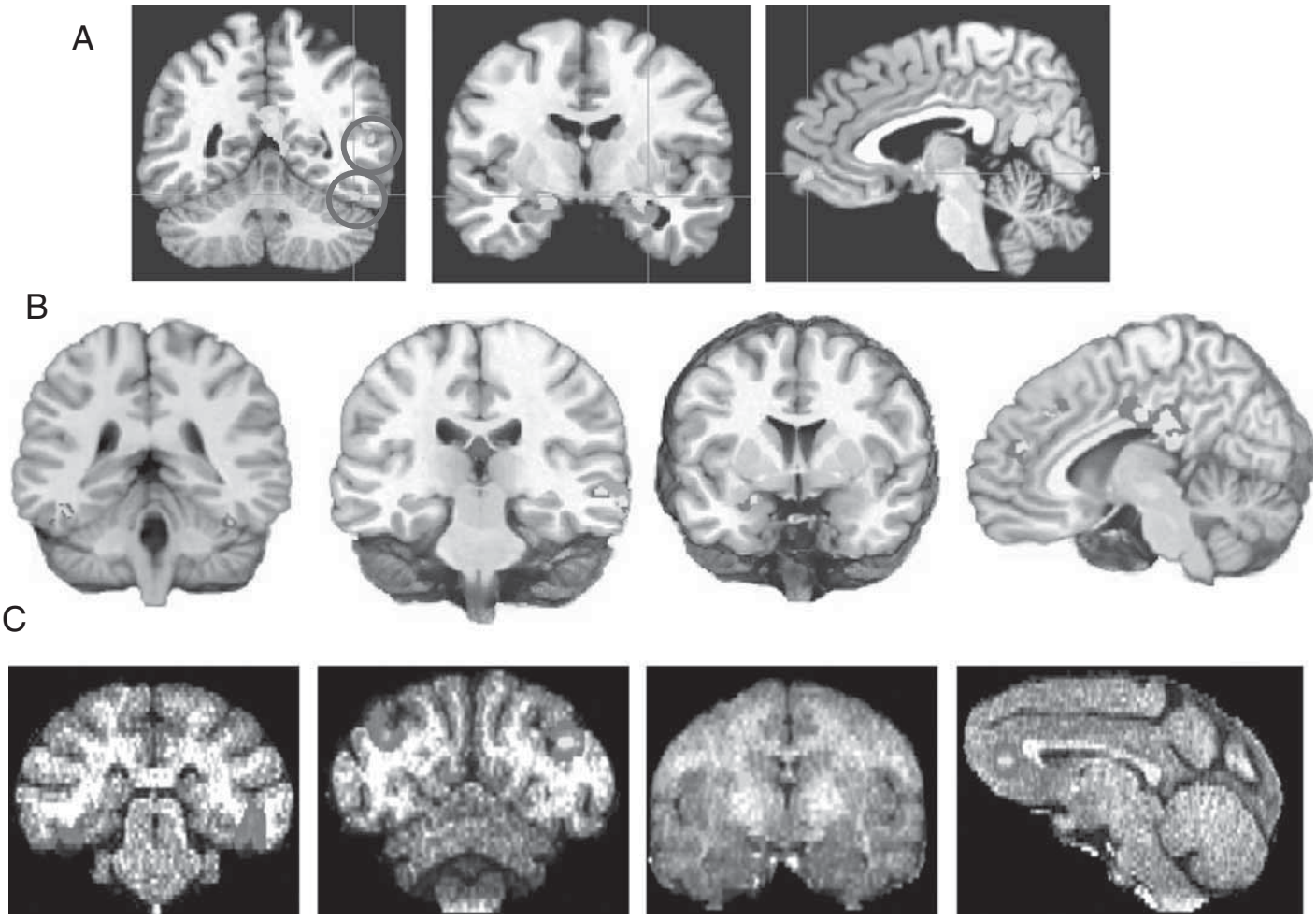


FIGURE 71.3 Correspondence across tasks and species in the location of the neural circuitry for perceiving and knowing about animate entities. (*A*) Regions shown in yellow were more active when subjects viewed photographs of faces relative to viewing photographs of common tools. Going from left to right, the first image shows a coronal slice through posterior cortex indicating the location of activity in the lateral portion of the right fusiform gyrus (lower red circle) and in the right pSTS (upper red circle). The next coronal image depicts bilateral activity in the amygdalae. The third image shows a sagittal section revealing activity in the medial prefrontal cortex and in the posterior cingulate/precuneus. (Unpublished data from our laboratory.) (*B*) Brain slices depicting conjunction of regions more active when subjects perceived simple shapes in motion as animate, relative to when they were judged to be inanimate, and when they imagined these stimuli as animate versus inanimate. Going from left to right, the first image is a coronal slice showing bilateral activity in the lateral fusiform gyrus. The next coronal slice shows the location of activity in the STS, the third depicts activity in the left amygdala, and the last shows activations located in the medial prefrontal and posterior cingulate cortices. (Adapted from Wheatley, Milleville, & Martin, 2007.) (*C*) Activity in the macaque brain when listening to species-specific calls. Shown are PET scans obtained from a single animal. Going from left to right, the first image shows a coronal slice through ventral regions TEO/TE, the next coronal slice shows activity in the STS, the third slice shows activation in the amygdala, and the fourth slice shows an activation located in Area 32 on the medial surface of the brain. (Adapted from Gil-da-Costa et al., 2004.) (See color plate 85.)

whenever an animate object is attended to. For example, as illustrated in figure 71.3*A*, simply viewing a face will produce activity throughout the entire circuit.

This characteristic of being activated whenever an object is viewed also holds for the nodes of the circuits underpinning perceiving and knowing about "tools." The current evidence suggests that the circuitry underpinning processing in this domain includes two regions in the posterior temporal lobe, one situated in the more medial extent of the posterior fusiform gyrus, the other located in the left posterior portion of the middle temporal gyrus. These regions have been linked to representing the visual form and visual motion associated with these objects (e.g., Beauchamp, Lee, Haxby, & Martin, 2002, 2003). The other two nodes in this circuit, both strongly lateralized to the left hemisphere, are in posterior parietal cortex (in the intraparietal sulcus and often also including a more anterior region in the inferior parietal lobule) and in ventral premotor cortices. These regions have

been linked to the representation of goal-directed action associated with an object's function (for recent reviews and discussion see Beauchamp & Martin, 2007; Frey, 2007; Lewis, 2006). As with the animacy circuitry discussed earlier, the circuitry underpinning perceiving and knowing about "tools" is engaged whenever these objects are viewed (e.g., Chao et al., 1999; Chao & Martin, 2000; Handy, Grafton, Sheroff, Ketay, & Gazzaniga, 2003) (see Mahon et al., 2007, for neuroimaging and neuropsychological evidence supporting the specificity of this circuitry for "tools" relative to other manmade, manipulable objects).

Activity in domain-specific neural circuitry transcends stimulus features

These findings suggest that the brain contains property-based, domain-specific neural circuits for perceiving and knowing about specific object categories. A case was made that one of these circuits developed for representing animate things, another for "tools." It was also suggested that these circuits are active whenever objects from these broad categories are perceived. However, this fact alone says nothing about the relationship between these circuits and conceptual processes. To make that link requires showing that activity in these circuits is associated with the *interpretation* of a stimulus, rather than its physical characteristics. There is now considerable evidence to support this claim. For example, the lateral region of the fusiform gyrus that has been linked to representing the visual form of animate entities responds to animate entities as represented by pictures and written names of animals (Chao et al., 1999; Devlin, Rushworth, & Mathews, 2005; Mechelli, Sartori, Orlandi, & Price, 2006; Okada et al., 2000; Price, Noppeney, Phillips, & Devlin, 2003; Rogers, Hocking, Mechelli, Patterson, & Price, 2005; Wheatley, Weisberg, Beauchamp, & Martin, 2005), human voices (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005), point-light displays of human bodies in motion (Beauchamp et al., 2003; Grossman & Blake, 2001, 2002; Peelen, Wiggett, & Downing, 2006), and humanlike stick figures (Peelen & Downing, 2005). In contrast, the more medial aspect of the fusiform associated with representing the visual form of "tools" has been reported in response to pictures and written names of tools (Chao et al.; Chao, Weisberg, & Martin, 2002; Devlin et al.; Mechelli et al.; Whatmough, Chertkow, Murtha, & Hanratty, 2002), the spoken names of tools (Noppeney, Price, Penny, & Friston, 2006), and point-light displays depicting tools in motion (Beauchamp et al.).

Perhaps even more convincingly, activity throughout the animacy circuit has even been observed when participants view abstract representations of social situations as illustrated by the interactions among simple geometric shapes in motion (Heider & Simmel, 1944). For example, the lateral fusiform gyrus responds to animations suggesting social interactions such as hide-and-seek (Schultz et al., 2003), mocking and bluffing (Castelli, Happe, Frith, & Frith, 2000; Castelli, Frith, Happe, & Frith, 2002), and sharing (Martin & Weisberg, 2003). These studies also reported activity in other nodes of the animacy circuit including pSTS, the amygdala, and ventromedial prefrontal cortices. In contrast, activity in the temporal lobe regions associated with the visual form and motion of "tools" (medial fusiform and left middle temporal gyrus, respectively) has been observed when animations composed of simple geometric shapes were interpreted as depicting mechanical interactions (Martin & Weisberg, 2003).

Wheatley and colleagues have recently provided even more compelling evidence that activity in these circuits is linked to the interpretation of a stimulus rather than its physical features (Wheatley, Milleville, & Martin, 2007). In that study, different background settings were used to bias the interpretation of a simple geometric shape in motion as depicting either an animate entity or an inanimate object. All the previously mentioned regions in the animacy circuit (lateral portion of the fusiform gyrus, STS, medial prefrontal cortex, posterior cingulate, amygdala) were active when the objects were interpreted as animate, relative to when that same form and motion were interpreted as depicting an inanimate object. Moreover, these regions were also active when subjects were asked to imagine the object they had previously seen based on viewing the backgrounds alone (figure 71.3*B*; see Wheatley et al., 2007, for details). Thus activation in this domain-specific, property-related circuit was not due to particular stimulus features, but rather appeared to be directly related to conceptual representation.

Several other studies have provided data to support this claim. Each of these studies used a learning paradigm to show that acquiring new information about novel objects changes the brain's response to those objects. Moreover, the locations of these responses were directly related to the type of information acquired. For example, Weisberg, van Turennout, and Martin (2007) asked subjects to perform a simple visual matching task on photographs of novel objects. After scanning, the subjects were given extensive training manually manipulating the objects to perform specific tool-like functional tasks. After training, the subjects were again scanned while performing the visual matching task. Comparison of the data collected prior to training with those collected after training revealed that experience using the objects as tools led to predictable changes in how these objects were now represented in the brain. Whereas prior to training visual matching of the novel objects elicited only broad activity in ventral occipitotemporal cortex, after training ventral temporal activity was largely restricted to the medial aspect of the fusiform gyrus, the same region previously implicated in representing the visual shape or form of

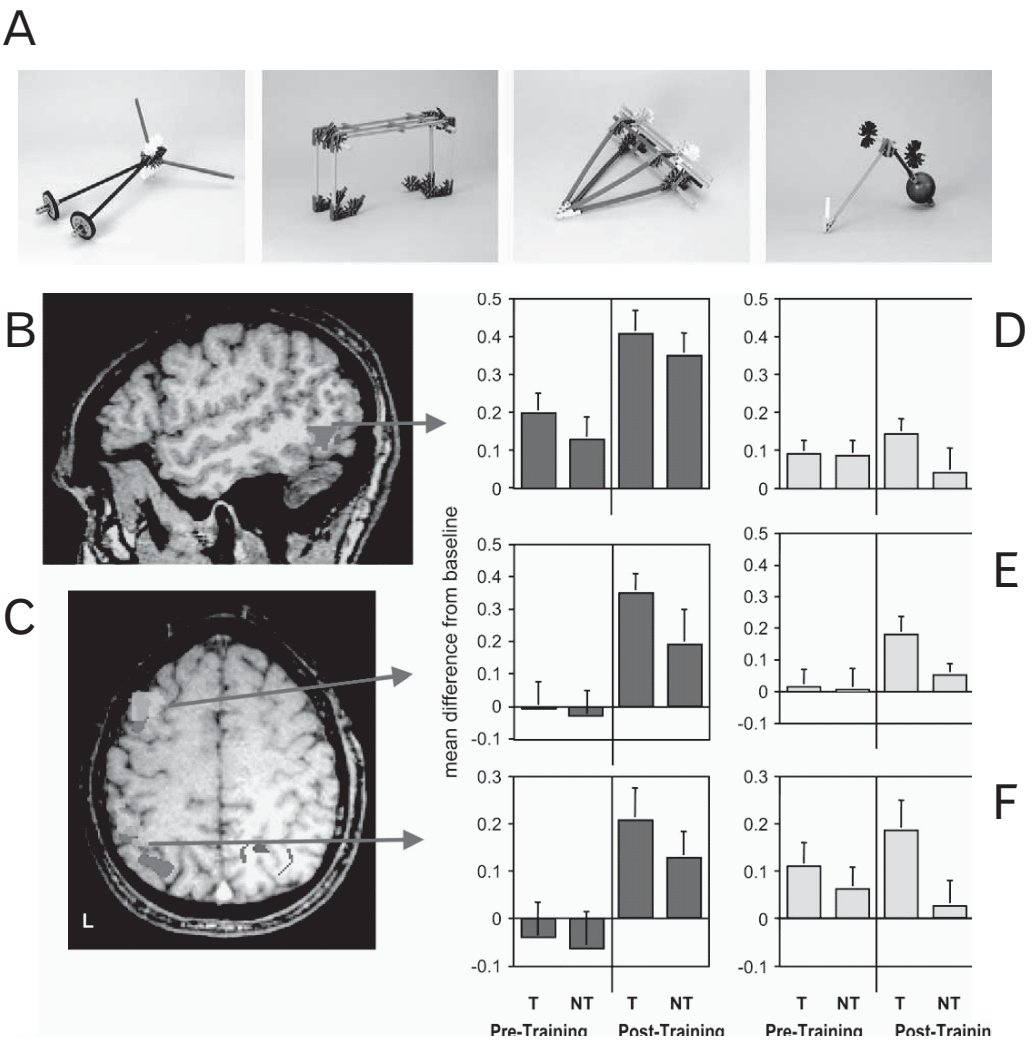


FIGURE 71.4 (*A*) Examples of novel objects designed to perform specific toollike functions. (*B*) Sagittal section showing the location of learning-related activity in the left middle temporal gyrus. Regions in red were more active after training than before training. Regions in yellow, which overlap with regions in red, were more active for trained (*T*) objects than for not-trained (*NT*) objects. (*C*) Axial section showing the location of learning-related activity in the left premotor/prefrontal cortex and intraparietal cortices. (*D*, *E*, *F*) Histograms showing the difference between novel-object-matching and scrambled-image-matching baseline task in the middle temporal gyrus, left premotor, and intraparietal regions, respectively. Red bars represent brain regions that showed increased activity for object matching after but not prior to training; yellow bars represent regions that demonstrated greater activity for trained objects than not-trained objects after but not prior to training. (Adapted from Weisberg, van Turennout, & Martin, 2007.) (See color plate 86.)

"tools." Similarly, new activations emerged after training in other regions of the circuitry associated with perceiving and knowing about "tools," including the left posterior region of the middle temporal gyrus (linked to nonbiological motion perception; Beauchamp et al., 2002, 2003), left intraparietal sulcus, and left premotor cortex (goal-directed manipulation related to object function) (figure 71.4).

Learning effects have also been observed for animate entities. It has been well documented that viewing point-light displays of human forms in motion elicits activity in lateral fusiform and pSTS (Beauchamp et al., 2003; Grossman & Blake, 2001, 2002). Grossman, Blake, and Kim (2004) trained subjects to perceive human forms in point-light displays that were embedded within visual noise. After training, not only were the subjects better at indicating when a human form was present in a noisy visual display, but they also exhibited greater fusiform and pSTS activity in response to detecting those forms, and the amount of activity in both regions was positively correlated with a subject's behavioral performance.

Finally, in addition to visual learning paradigms, it has been demonstrated that a verbal learning procedure can be used to demonstrate the development of property-based circuitry (T. James & Gauthier, 2003). Prior to scanning, subjects learned verbally presented information about the auditory and motor-related properties of different families

of novel animate-like entities ("greebles"). For example, subjects were trained that a particular family of greebles were associated with an auditory property (e.g., roars or squeaks), whereas other types of greebles had action properties (e.g., hops or jumps). After training, subjects underwent fMRI while performing a visual matching task that did not require retrieval of these learned associations. The results showed that viewing greebles associated with auditory properties produced activity in auditory cortex (as defined by an auditory functional localizer) and viewing greebles associated with action properties produced activity in the biological-motion-sensitive region of the pSTS (as localized by moving point-light displays). These findings, along with the findings of Weisberg et al. (2007), demonstrate that experience with novel objects leads to the development of activity in domain-specific property circuits. Simply seeing an object from the training set elicited activity in specific regions of the previously described circuits, even though that information was not necessary for successfully performing the task and not present in the stimuli. Subjects learned that, for example, a particular object was associated with a particular type of movement (e.g., hopping). Having acquired that knowledge, a region in pSTS that is active when viewing biological motion became active when that object was viewed, even though the subject's task did not require retrieving that information. The posterior region of the STS was activated automatically when the object was seen again.

A mechanism that allows us to quickly and effortlessly form inferences about objects in the world has obvious survival value. As a result we would expect that the ability to infer properties would be preserved across primate species. Recent evidence suggests that this may be the case with regard to the circuitry supporting perception of animate entities. Using PET to study perception of species-specific calls in the macaque, Gil-da-Costa and colleagues (2004) showed that the calls elicited activity in area TE/TEO, a presumed monkey homologue of human fusiform gyrus, and in the pSTS, relative to acoustically similar controls (figure 71.3*C*). In addition, calls known to carry emotional connotations activated the amygdala and medial prefrontal cortices over and above calls presumed to connote more neutral associations (see Gil-da-Costa et al. for details). While strong claims cannot be made about the meaning of these calls for the macaques, it should be safe to conclude, at the very least, that the calls were interpreted as indicating the presence of another monkey. Thus, as with humans, when monkeys process information about animate entities, activation occurs across a distributed circuit. The nodes of this circuitry are presumed to represent the salient properties of those entities, including what they look like and how they move, even when those properties are not present in the stimulus, and therefore must be inferred.

Additional architectural considerations: The role of the anterior regions of the temporal lobes

Clearly, these circuits do not operate in isolation. For one thing, information must be selected and retrieved, and much work has established that the left inferior prefrontal cortex plays a prominent role in performing these functions (Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; for review see Thompson-Schill, Bedny, & Goldberg, 2005). Object-property information must also be integrated, and this requirement raises a form of the binding problem on the level of conceptual representation. One potential mechanism for achieving an integration of information stored in different locations is through their interaction. In that scenario, each node would represent the information it was specialized for, as well as reflecting or re-representing other types of information stored elsewhere (see Konen & Kastner, 2008, and Schwarzlose, Swisher, Dang, & Kanwisher, 2008, for neuroimaging data consistent with this view). Another possibility is that information from all circuits is integrated in a specific region. Several candidates have been proposed for this "hub" architecture, including posterior regions of left lateral temporal cortex (Hickok & Poeppel, 2000), left prefrontal cortex (reviewed in Thompson-Schill et al.), and thalamus (Kraut et al., 2002).

More recently, a highly influential version of a hub architecture has been proposed that locates this mechanism in the most anterior portion of the temporal lobes (Lambon Ralph, Lowe, & Rogers, 2007; McClelland & Rogers, 2003; Patterson et al., 2007; also see McClelland et al., chapter 72 in this volume). As argued by Patterson and colleagues (2007) in order to operate in the service of semantic cognition, property-based circuits require that all stored information about objects be integrated at a single location (Patterson et al.). Under this view, a central hub is needed because a distributed architecture alone cannot account for one of the central defining characteristics of a conceptual system; the ability to generalize across exemplars belonging to the same category (e.g., telephone) when the specific exemplars in this category can have very different physical features (desk phones, cellular phones) (see Patterson et al. for details of this argument). Moreover, according to this view, the ability to generalize requires amodal conceptual representations, as opposed to the modality-based representations described here. Amodal representations require a central hub (Patterson et al.; also see McClelland & Rogers, 2003).

A hub of this type may in fact be necessary on computational grounds, and that possibility will not be disputed here. It should be stressed, however, that arguments about the need for a conceptual hub and the physical location of that hub in the brain are independent. It is this later claim, specifically the claim that the hub is located in the most anterior part of the temporal lobes, that I will address here.

The anterior temporal lobes include a number of distinct anatomical divisions, including the temporal pole, amygdala, and entorhinal and perirhinal cortices, as well as the anterior extents of the fusiform, inferior, middle, and superior temporal gyri. Therefore it should not be surprising that there are currently several different but non-mutually-exclusive views of anterior temporal lobe function. One view, and one that has the most support from neuroimaging and neuropsychological investigation, is that the anterior regions of the temporal lobe are involved in social and emotional processing (see Olson et al., 2007, for a recent review). Another view, also supported by neuroimaging (Gorno-Tempini & Price, 2001; Grabowski et al., 2001) and neuropsychological investigation (e.g., Tranel, Damasio, & Damasio, 1997; Damasio 1989), is that the anterior temporal lobes are involved in representing unique entities (i.e., famous people and places). A third view is that anterior temporal regions play a role in modulating access to distributed modality-specific information stored elsewhere, but are not involved in integrating this information (Martin & Chao, 2001). Finally, a fourth position is that the anterior temporal lobes are the location of the conceptual hub. Support for this claim comes primarily from study of patients with semantic dementia, a progressive disorder that is associated with pathology with a proclivity for attacking the temporal lobes, especially the more anterior portion where the damage often appears to originate (McClelland & Rogers, 2003; Patterson et al., 2007).

Several points are in order. First, studies using voxel-based morphometry to measure the extent of atrophy associated with semantic dementia indicate that pathology in these patients is not limited to the anterior temporal lobes. Rather, these studies uniformly show that the pathology often extends to the more posterior regions of the temporal lobes engaged in many of the neuroimaging studies reviewed previously. Moreover, the semantic deficits in these patients are nearly as strongly related to atrophy in posterior temporal cortex as with atrophy in anterior temporal cortex (G. Williams, Nestor, & Hodges, 2005). Semantic dementia is a progressive disorder. As symptoms increase in severity, pathology gets more widespread throughout the temporal lobes. Thus the discrepancy between the findings with semantic dementia patients and the neuroimaging literature may not be nearly as strong as some have suggested (Patterson et al., 2007). It is probably also noteworthy that these patients often have pathology outside the temporal lobes, most prominently in frontal cortex. Thus it is not at all clear that the devastating impairments in semantic cognition that characterize these patients can be attributed solely to anterior temporal lobe pathology (Lambon Ralph et al., 2007). The available evidence suggests that the anterior regions of the temporal lobes likely support multiple functions. It is more than likely that one of these functions involves conceptual and semantic processing. The exact nature of this role, however, and, in particular, whether the anterior temporal lobes are necessary for creating amodal representations, remains to be determined.

Summary and concluding comments

The evidence discussed in this chapter indicates that the information about salient object properties—such as how they look, move, and are used, along with our affective associations to them—is stored in the neural systems that support perceiving, acting, and feeling. It is in this sense that conceptual knowledge is argued to be grounded and embodied. The evidence further suggests that this information is not stored in every part of our sensory and motor systems. The circuits for sensing, perceiving, and knowing are partially, not fully, overlapping.

These architectural constraints, however, say nothing about the nature or format of this information. It has been assumed that information stored in discrete regions of the fusiform gyrus represents the visual form of objects. This assumption has been made because this region is part of the ventral visual object-processing stream known to underpin object identification. Yet it appears that these same regions respond in a categorical manner in the blind when palpating objects (Pietrini et al., 2004). This finding is consistent with the idea that this region codes from object shape or form, but it also suggests that the way shape is represented may be quite abstract. Information about object shape may be stored in the ventral stream, even when that shape information was obtained through a different modality, in this case touch rather than vision. This finding, in turn, challenges us to specify the sense in which the information grounded in perceptual and action systems should be considered modality specific or sensory or motor in nature.

The evidence reviewed here also suggests that object-property-based information is organized into broadly defined domain-specific circuits. These circuits appear to be remarkably stable in the sense that the spatial arrangement among their defining nodes seem to be consistent from one individual to another. This stability is most apparent when considering the spatial arrangement of regions in ventral temporal cortex purported to support identification of words (McCandliss, Cohen, & Dehaene, 2003), faces (Yovel & Kanwisher, 2004), animals (Chao et al., 1999), tools (Chao et al., 2002), and environmental scenes (Epstein, 2008). Although discussion of this important issue is outside the scope of this chapter, several suggestions have been offered to explain this fact (Op de Beeck, Haushoffer, & Kanwisher, 2008; Martin, 2006; Mahon et al., 2007).

The evidence also suggests that activation of these circuits is dependent on how a stimulus or event is interpreted, not on the physical features of the stimuli impinging on our

senses. These findings indicated that while different regions of the cortex are specialized for processing and storing information about specific properties (e.g., biological motion), these same regions can be reactivated in top-down fashion based the interpretation applied to a stimulus, even when that critical property is not physically present (Wheatley et al., 2007).

This type of finding raises important questions about the function played by these activated regions. In the Wheatley and colleagues (2007) study, the task was simply to indicate whether the depicted object represented an animate thing. Nevertheless, animacy identification led to activation of a suite of regions that, based on other data, support a range of complex higher-order social processes (e.g., theory of mind, making self-other similarity judgments; Mitchell, 2008). In fact, as illustrated in figure 71.3*A*, this entire circuit becomes active when simply viewing photographs of the human face. Clearly, activation in ventromedial prefrontal cortex or in the amygdala is not necessary to perceive faces. Thus a major challenge for future studies is to specify what role these activations play in these tasks. One possibility, and I believe the most likely explanation, is that these activations reflect the automatic generation of inferences that are a central part of what we mean by a conceptual representation. In this sense, these activations may serve to prime the conceptual system for future action. That is, they are predictive of future events. Seeing a hammer activates the dorsal stream because hammers are objects likely to be grasped and used to perform some function. Seeing other individuals activates a broad circuit of regions so we are prepared to interpret their state of mind and actions. Our ability to sort out the role played by these regions in the context of different tasks will require investigations that combine functional neuroimaging and lesion approaches. Investigations of this type have just begun, but they have already yielded tantalizing clues (Calder, Keane, Manes, Antoun, & Young, 2000; Mahon et al., 2007).

NOTE

1. It should be stressed that the overlap observed in this and in other studies of this type does not necessarily mean that the *same neurons* are involved in both perceiving and knowing. Support for that claim would require single-unit recordings from the human brain. Functional neuroimaging evidence consistent with this claim could be obtained by showing that the amplitude of the BOLD signal in a region of cortex was reduced when verbally retrieving information about a property, for example, color, following activity produced by viewing that color (i.e., by showing an across-task repetition suppression effect. For a discussion of the logic behind this approach see Grill-Spector & Malach, 2001; Henson, 2003). Nevertheless, the embodiment view proposed here does not require that perceiving and knowing be coded in the same neurons. It does, however, require that these processes be carried out in the same brain region, strictly defined. For example, the embodied view would hold if the neurons involved in visual perception and those involved in information storage were found to be interdigitated in the same tightly constrained space.

REFERENCES

Adolphs, R. (2001). The neurobiology of social cognition. *Curr. Opin. Neurobiol.*, *11*, 231–239.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Badre D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*, 907–918.

Barsalou, L. W. (1989). Ad hoc categories. *Mem. Cogn.*, *11*, 211–227.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.*, *22*, 637–660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617–645.

Beauchamp, M. S., & Martin, A. (2007). Grounding object concepts in perception and action: Evidence from fMRI studies of tools. *Cortex*, *43*, 461–468.

Beauchamp, M. S., Haxby, H. V., Jennings, J. E., & DeYoe, E. A. (1999). An fMRI version of the Farnsworth-Munsell 100-Hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cereb. Cortex*, *9*, 257–263.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*,149–159.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). fMRI responses to video and pointlight displays of moving humans and manipulable objects. *J. Cogn. Neurosci.*, *15*, 991–1001.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.*, *25*, 151–188.

Broadbent, W. H. (1878). A case of peculiar affection of speech with commentary. *Brain*, *1*, 484–503.

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nat. Neurosci*, *3*, 1077–1078.

Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cogn. Neuropsychol.*, *20*, 213–261.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *J. Cogn. Neurosci.*, *10*, 1–34.

Castelli, F., Frith, C., Happe, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental stats to animated shapes. *Brain*, *125*, 1839–1849.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*, 314–325.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat. Neurosci.*, *2*, 913–919.

Chao, L. L., & Martin, A. (1999). Cortical representation of perception, naming, and knowledge of color. *J. Cogn. Neurosci.*, *11*, 25–35.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, *12*, 478–484.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cereb. Cortex*, *12*, 545–551.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*, 25–62.

Devlin, J. T., Rushworth, M. F. S., & Matthews, P. M. (2005). Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia*, *43*, 69–74.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.*, *12*, 388–396.

Etcoff, N. L., Freeman, R., & Cave, K. R. (1991). Can we lose memories of faces—Content specificity and awareness in a prosopagnosic. *J. Cogn. Neurosci.*, *3*, 25–41.

Freud, S. (1891/1953). *On aphasia* (E Stengel, Trans.). New York: International University Press.

Frey, S. H. (2007). What puts the how in where? Tool use and the divided visual streams hypothesis. *Cortex*, *43*, 368–375.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.*, *22*, 455–479.

Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M. D., Carson, R. E., Herscovitch, P., & Martin, A. (2004). Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual and affective processing systems in the macaque. *Proc. Natl. Acad. Sci. USA*, *101*, 17516–17521.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *J. Neurosci.*, *26*, 4917–4921.

Gorno-Tempini, M. L., & Price, C. J. (2001). Identification of famous faces and buildings—A functional neuroimaging study of semantically unique items. *Brain*, *124*, 2087–2097.

Grabowski, T. J., Damasio, H., Tranel, D., Ponto, L. L. B., Hichwa, R. D., & Damasio, A. R. (2001). A role for left temporal pole in the retrieval of words for unique entities. *Hum. Brain Mapping*, *13*, 199–212.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition—As soon as you know it is there, you know what it is. *Psychol. Sci.*, *16*, 152–160.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychol. (Amst.)*, *107*, 293–321.

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Res.*, *41*, 1475–1482.

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, *35*, 1167–1175.

Grossman, E. D., Blake, R., & Kim, C.-Y. (2004). Learning to see biological motion: Brain activity parallels behavior. *J. Cogn. Neurosci.*, *16*, 1669–1679.

Handy, T. C., Grafton, S. T., Shroff, N. M., Ketay, S., & Gazzaniga, M. S. (2003). Graspable objects grab attention when the potential for action is recognized. *Nat. Neurosci.*, *6*, 421–427.

Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.*, *4*, 223–233.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *Am. J. Psychol.*, *57*, 243–249.

Henson, R. N. (2003). Neuroimaging studies of priming. *Prog. Neurobiol.*, *70*, 53–81.

Hickok, G. & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.*, *4*, 131–138.

James, T. W., & Gauthier, I. (2003). Auditory and action semantic features activate sensory-specific perceptual brain regions. *Curr. Biol.*, *13*, 1792–1796.

James, W. (1890/1950). *The principles of psychology*. New York: Dover.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proc. Natl. Acad. Sci. USA*, *101*, 3310–3315.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nat. Neurosci.*, *11*, 224–231.

Kraut, M. A., Kremen, S., Segal, J. B., Calhoun, V., Moo, L., & Hart, J. (2002). Object activation from features in the semantic system. *J. Cogn. Neurosci*, *14*, 24–36.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. New York: Basic Books.

Lambon Ralph, M. A., Lowe, C., & Rogers, T. T. (2007). Neural basis of category-specific semantic deficits for living things: Evidence from semantic dementia, HSVE and a neural network model. *Brain*, *130*,1127–1137.

Lewis, J. W. (2006). Cortical networks related to human use of tools. *Neuroscientist*, *12*, 211–231.

Lissauer, H. (1890/1988). A case of visual agnosia with a contribution to theory (M. Jackson, Trans.). *Cogn. Neuropsychol.*, *5*, 157–192.

Mahon, B. Z., & Caramazza, A. (2005). The orchestration of the sensory-motor systems: Clues from neuropsychology. *Cogn. Neuropsychol.*, *22*, 480–494.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris*, *102*, 59–70.

Mahon, B. Z., Milleville, S. C., Negri, G. A. L., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron*, *55*, 507–520.

Martin, A. (1998). The organization of semantic knowledge and the origin of words in the brain. In N. Jablonski & L. Aiello (Eds.), *The origins and diversification of language* (pp. 69–98). San Francisco: California Academy of Sciences.

Martin, A. (2001). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 153–186). Cambridge, MA: MIT Press.

Martin, A. (2006). Shades of Dejerine: Forging a causal link between the visual word form area and reading. *Neuron*, *50*, 173–175.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25–45.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Curr. Opin. Neurobiol.*, *11*, 194–201.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated

with knowledge of color and knowledge of action. *Science*, *270*, 102–105.

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cogn. Neuropsychol.*, *20*, 575–587.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.*, *7*, 293–299.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nat. Rev. Neurosci.*, *4*, 310–322.

Mechelli, A., Sartori, G., Orlandi, P., & Price, C. J. (2006). Semantic relevance explains category effects in medial fusiform gyri. *NeuroImage*, *30*, 992–1002.

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*, 1013–1052.

Miller, E. K., Nieder, A., Freedman, D. J., Wallis, J. D. (2003). Neural correlates of categories and concepts. *Curr. Opin. Neurobiol.*, *13*, 198–203.

Mitchell, J. P. (2008). Contributions of functional neuroimaging to the study of social cognition. *Curr. Dir. Psychol. Sci.*, *17*, 142–146.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *NeuroImage*, *28*, 757–762.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *4*, 655–663.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Noppeney, U., Josephs, O., Kiebel, S., Friston, K. J., & Price, C. J. (2005). Action selectivity in parietal and temporal cortex. *Cogn. Brain Res.*, *25*, 641–649.

Noppeney, U., Price, C. J., Penny, W. D., Friston, K. J. (2006). Two distinct neural mechanisms for category-selective responses *Cereb. Cortex*, *16*, 437–445.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychol. Rev.*, *108*, 483–522.

Okada, T., Tanaka, S., Nakai, T., Nishizawa, S., Inui, T., Sadato, N., Yonekura, Y., & Konishi, J. (2000). Naming of animals and tools: A functional magnetic resonance imaging study of categorical differences in the human brain areas commonly used for naming visually presented objects. *Neurosci. Lett.*, *296*, 33–36.

Olson, I. R., Ploaker, A., & Ezzyat, Y. (2007). The enigmatic temporal pole: A review of findings on social and emotional processing. *Brain*, *130*, 1718–1731.

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nat. Rev. Neurosci.*, *9*, 123–135.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.*, *8*, 976–987.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *J. Neurophysiol., 93*, 603–608.

Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron*, *49*, 815–822.

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proc. Natl. Acad. Sci. USA*, *99*, 11458–11463.

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annu. Rev. Psychol.*, *57*, 27–53.

Phelps, E. A., & Ledoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*, 175–187.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W.-H. C., Cohen, L., Guazzelli, M., & Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proc. Natl. Acad. Sci. USA*, *101*, 5658–5663.

Price, C. J., Noppeney, U., Phillips, J., & Devlin, J. T. (2003). How is the fusiform gyrus related to category-specificity? *Cogn. Neuropsychol.*, *20*, 561–574.

Riddoch, M. J., Johnston, R. A., Bracewell, R. M., Boutsen, L., & Humphreys, G. W. (2008). Are faces special? A case of pure prosopagnosia. *Cogn. Neuropsychol.*, *25*, 3–26.

Rogers, T. T., Hocking, J., Mechelli, A., Patterson, K., & Price, C. J. (2005). Fusiform activation to animals is driven by the process, not the stimulus. *J. Cogn. Neurosci.*, *17*, 434–445.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cogn. Psychol.*, *8*, 382–439.

Saxe, R. (2006). Uniquely human social cognition. *Curr. Opin. Neurobiol.*, *16*, 235–239.

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., & Skudlarski, P. (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *358*, 415–427.

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *J. Neurosci.*, *25*(47), 11055–11059.

Schwarzlose, R. F., Swisher, J. D., Dang, S. B., Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proc. Natl. Acad. Sci. USA*, *105*, 4447–4452.

Shuren, J. E., Brott, T. G., Schefft, B. K., & Houston, W. (1996). Preserved color imagery in an achromatopsic. *Neuropsychologia*, *34*, 485–489.

Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cereb. Cortex*, *15*, 1602–1608.

Simmons, W. K., Ramjee, V., McRae, K., Martin, A., & Barsalou, L. W. (2007). fMRI evidence for an overlap in the neural bases of color perception and color knowledge. *Neuropsychologia*, *45*, 2802–2810.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: Inferring "how" from "where." *Neuropsychologia*, *41*, 280–292.

Thompson-Schill, S. L., Bedny, M., Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Curr. Opin. Neurobiol.*, *15*, 219–224.

Thompson-Schill, S. L., Kan, I. P., & Oliver, R. T. (2006). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (2nd ed., pp. 149–190). Cambridge, MA: MIT Press.

Tranel, D., Damasio, H., & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*, *35*, 1319–1327.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat. Neurosci.*, *6*, 624–631.

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *J. Cogn. Neurosci.*, *17*, 367–376.

Weisberg, J., van Turennout, M., & Martin, A. (2007). A neural system for learning about object function. *Cereb. Cortex*, *17*, 513–521.

Whatmough, C., Chertkow, H., Murtha, S., & Hanratty, K. (2002). Dissociable brain regions process object meaning and object structure during picture naming. *Neuropsychologia*, *40*, 174–186.

Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror system. *Psychol. Sci.*, *18*, 469–474.

Wheatley, T., Weisberg, J., Beauchamp, M. S., & Martin, A. (2005). Automatic priming of semantically related words reduces activity in the fusiform gyrus. *J. Cogn. Neurosci.*, *17*, 1871–1885.

Williams, G. B., Nestor, P. J., & Hodges, J. R. (2005). Neural correlates of semantic and behavioural deficits in frontotemporal dementia. *NeuroImage*, *24*, 1042–1051.

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *J. Neurosci.*, *24*, 2898–2904.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.*, *9*(4), 625–636.

Yang, J. J., Bellgowan, P. S. F., & Martin, A. (2008). Interaction of category and affect in the human amygdala. Under review.

Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron*, *44*, 889–898.

Zeki, S. (1990). A century of cerebral achromatopsia. *Brain*, *113*, 1721–1777.

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.*, *11*, 641–649.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *J. Exp. Psychol. Gen.*, *135*, 1–11.

72 Semantic Cognition: Its Nature, Its Development, and Its Neural Basis

JAMES L. McCLELLAND, TIMOTHY T. ROGERS, KARALYN PATTERSON, KATIA DILKINA, AND MATTHEW LAMBON RALPH

ABSTRACT We consider parallels in the development of conceptual knowledge and its disintegration in a neurological disorder. During development, children's concepts become more differentiated. Developing children make striking errors, attributing properties to objects that they do not have and over-extending frequent category labels to other objects. Patients with semantic dementia (SD), a progressive condition affecting the anterior temporal lobes, show a loss of differentiation of concepts. Like developing children, they attribute properties to objects that they do not have and over-extend frequent category labels. Such patients may draw four legs on a duck, or call an elephant a horse. We present a theoretical framework and computational models that explain these parallel phenomena. We also consider the disintegration of lexical knowledge, which parallels the deterioration of semantics in SD patients. We present a model that simulates these findings within a single system that processes and represents both lexical and semantic information.

Interest in the nature of conceptual knowledge extends back at least to the ancient Greek philosophers. In recent years, there has been a wide range of different approaches to understanding the nature of conceptual knowledge, its development, and its neural basis. In most other work, however, these issues are not all treated together. Instead, workers in philosophy, adult experimental psychology, child development, and cognitive neuroscience have pursued related questions in relative ignorance of each other's efforts. Even within cognitive neuroscience, there has been until recently a relative separation between approaches taken by neuropsychologists, who study the effects of brain disease on cognition in patients, and researchers who study the neural basis of conceptual knowledge in neurologically intact populations, using functional imaging and related methods.

We have sought to develop an integrative perspective on these matters. Our effort is facilitated by our theoretical framework, which lends itself to implementation in computational models that capture both the ability to learn gradually from experience, in order to model development, and the tendency to degrade in a graded fashion, in order to capture the partial nature of the deficits resulting from brain injury. We begin with an overview of the theoretical framework and its application to three developmental phenomena. We then show how the framework also addresses parallel phenomena that arise in the striking neuropsychological condition called semantic dementia. We review evidence that the disorder affects knowledge of words as well as knowledge of things, motivating an extension of our theory in which knowledge of things and words is fully integrated, contra most other approaches. A final section describes imaging and magnetic stimulation studies in normals that test predictions arising from the theory; this section also considers evidence from disorders other than semantic dementia that indicates how the theory might be extended to address the flexible use of semantic knowledge in complex task situations.

The PDP framework and its application to development

The parallel distributed processing (PDP) approach (Rumelhart, McClelland, & the PDP Research Group, 1986) provides the starting point for our theory of semantic cognition. The fundamental tenets of the approach are as follows:

• Cognitive activities emerge from the interactions of large numbers of simple processing units and are distributed over populations of such units both within and across brain areas.

• Active representations in this framework—the representation someone may have when, for example, he brings to

JAMES L. McCLELLAND Department of Psychology, Stanford University, Stanford, California
TIMOTHY T. ROGERS Department of Psychology, University of Wisconsin, Madison, Wisconsin
KARALYN PATTERSON MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom
KATIA DILKINA Department of Psychology, Stanford University, Stanford, California
MATTHEW LAMBON RALPH Department of Psychology, University of Manchester, Manchester, United Kingdom

mind a particular dog, who may be greeting him by jumping up and down, barking, and licking at his face—are likewise distributed, involving activity of many contributing units in each of many disparate brain areas that contain neurons representing the shape, color, odor, movement, and sounds made by the dog being imagined.

• The ability of one kind of information (say the sight of a particular dog, or its bark, or simply the word "dog") to bring this information to mind depends on knowledge stored in the patterns of strengths or "weights" on the connections among the participating neurons.

• The patterns of connection weights are gradually acquired through experience. This process takes place over developmental time (i.e., years), gradually affecting the details of the representations we are able to bring to mind from particular inputs, and thus shaping gradual change in our behavior in cognitive tasks.

• The units and the connections are the substrate we use to understand not only normal cognitive functions and their development, but also the effects of brain damage and brain diseases on these functions. In particular, we assume that damage or disease results in the loss or disruption of the units and/or connections.

Within this theory, the effects of experience on connection weights explain many aspects of conceptual development; and the effects of damage and disease on the units and connections explain the disintegration of these functions in conceptual disorders such as semantic dementia.

It should be noted that our theory posits two complementary learning systems, only one of which—the neocortical one—is the focus here. We also propose that there is a second, fast-learning system based in the medial temporal lobes, allowing rapid learning of new, arbitrary information (McClelland, McNaughton, & O'Reilly, 1995). Normally, new semantic learning depends on both systems working together, but semantic knowledge is thought to gradually become independent of the fast-learning system, as evidenced by patients with profound amnesia for new information who nevertheless retain their semantic knowledge and abilities (Squire, 1992).

Some Phenomena in Conceptual Development As children develop, their conceptual knowledge gradually changes, in a way that appears to reflect aspects of experience. Three key features of this process are the following:

• Progressive differentiation of conceptual knowledge in development

• A tendency to overgeneralize names of frequently occurring objects

• A tendency to produce "illusory correlations" in attributing properties to objects

Differentiation in development has been explored in several contexts. In one investigation, Keil (1979) asked children what kinds of attributions or "predications" could apply to particular objects. For example, he asked children if it was "ok" or "silly" to say that something (say, a movie or a chair) "is sorry" or "is an hour long." Using their responses, he constructed, for each of many children at each of several ages, a "predicability tree" like the ones shown in figure 72.1. Here we see trees for representative children at progressively older ages. Clearly, these trees indicate conceptual differentiation. As children grow older, they cease to lump together concepts that older children (and adults) pull apart. Similar conclusions arise from work on infants using nonverbal methods (e.g., Mandler & McDonough, 1993; Pauen, 2002).

Overgeneralization of frequently encountered names is also well documented. Such behavior is striking since it often represents an error that first emerges and then subsequently disappears—a classic "U-shaped" trend in development. As Mervis (1987) has investigated in detail, one common case of such overgeneralization is the extension of the word "dog" as a name for a wide variety of different animals, particularly other four-legged animals, by children in the early childhood years.

The developmental phenomenon of "illusory correlations" has been treated by many investigators as a sign that young children have acquired domain-specific causal theories that lead them to overapply properties attested in these theories to other objects (Keil, 1991). One such illusory correlation was observed by Gelman and Williams (1998). They showed pictures of objects to young children and asked if the objects could go up and down a hill by themselves. Children usually answered yes if the pictured object appeared to be an animal. When asked to explain their answers, children often attributed feet to the pictured animals, even in cases where no feet were in evidence in the pictures. While in some cases (e.g., an animal called an echidna, which appeared as a small furry ball) such attributions were justified; in other cases (e.g., a snake) they would clearly be "illusory correlations"—perceived correlations of movement properties with physical properties of objects that are often, but not always, valid. Such overattributions are not by any means restricted to young children, but also occur in adult cognition. For example, people in the United States have a tendency to perceive an innocuous object as a weapon when it is in the hands of an African American (Eberhart, Goff, Purdie, & Davies, 2004).

Application of the Theory to Development: The Rumelhart Model Our application of our theory to conceptual development grew out of earlier work by Hinton (1981, 1989) and Rumelhart (1990). Rumelhart wished to

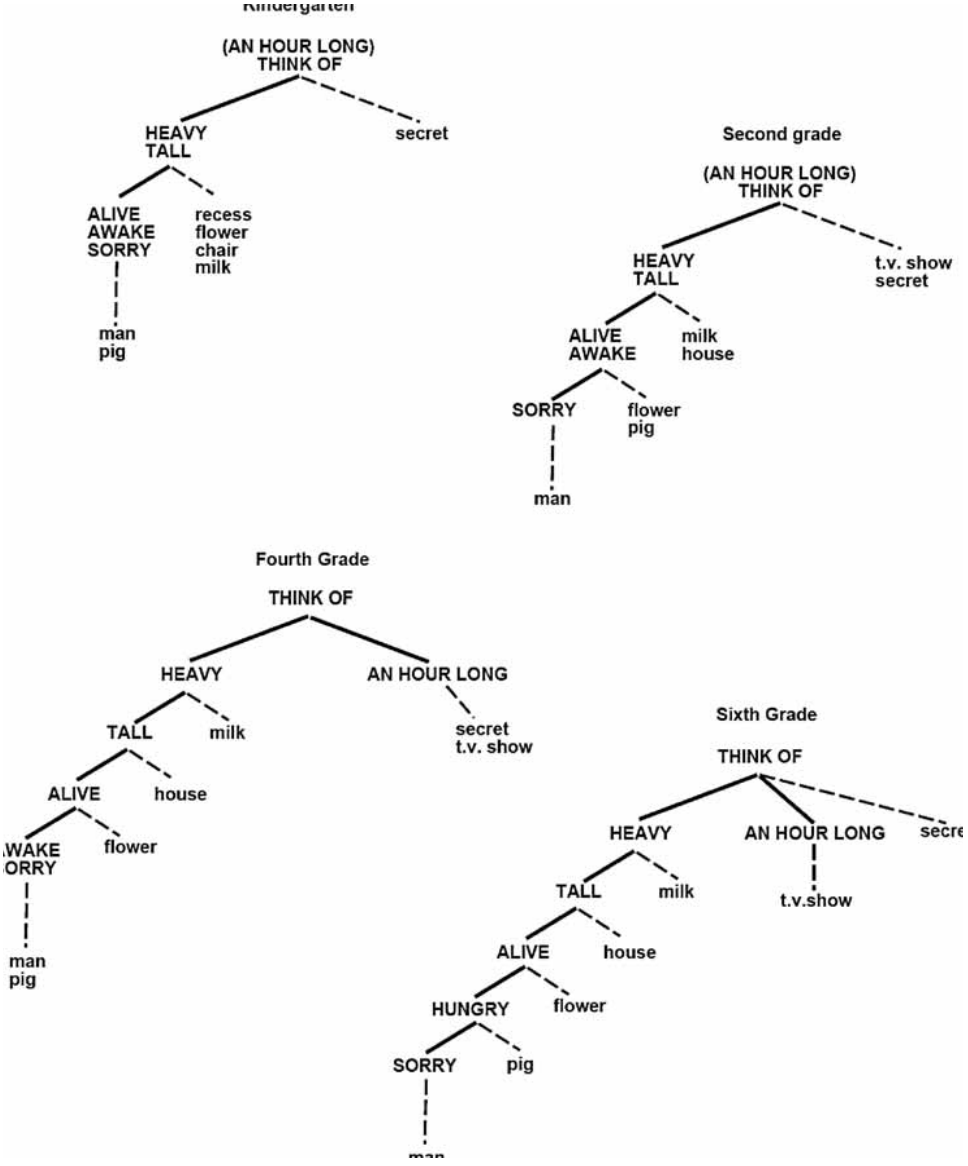


FIGURE 72.1 Examples of the predicability trees constructed by Keil (1979), from four individual children in different age groups. The trees indicate which of several predicate terms are accepted by the individual children as being applicable to various concepts; concepts accepting the same set of predicates are grouped together at the same branch of the tree. (Reprinted with permission from Rogers & McClelland, 2004, figure 1.3, p. 10, based on appendices A3, A17, A37, and A54 from Keil, 1979, pp. 181, 183, 185, and 187.)

articulate an alternative to prior ways of thinking about conceptual knowledge and chose the domain of living things as his example. Instead of storing knowledge explicitly as a network of linked propositions, as shown in figure 72.2, he proposed it might be stored in connections among simple processing units, as in the network shown in figure 72.3. In this network, called the *Rumelhart network*, the knowledge, say, that a robin can grow, can move, and can fly is not stored in explicit propositions, but instead arises from a pattern-completion process. We query the network to tell us what a robin can do by activating units for "robin" and "can" on the input side of the network, and propagating activation forward through intermediate layers to the output layer. The network is simplified in many ways relative to our overall theory, but it provides a useful ground for explaining the patterns we have discussed that arise in development.

When it is first initialized, the network's connection weights are both small and random, so that a query produces very neutral and undifferentiated patterns of activation at all levels of the network forward from the item and context inputs. However, the network is trained with repeated exposure to the information in figure 72.2. Each input consists of

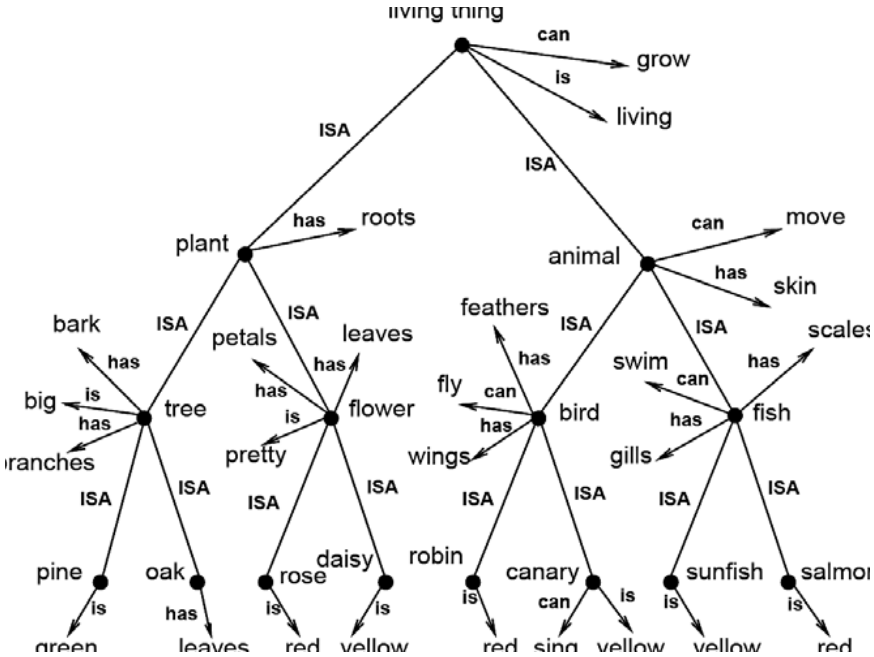


FIGURE 72.2 A taxonomic hierarchy of the type used by Quillian (1968) in his propositional model of the organization of knowledge in memory. ISA links may be followed up the tree to infer properties not explicitly connected to items below. (Reprinted with permission from Rogers & McClelland, 2004, figure 1.2, p. 6.)

an item (one of the eight items at the bottom of the figure) in each of four relational contexts (called "is" for appearance properties, "ISA" for the categories to which it belongs, "can" for things it can do, and "has" for specifying its parts). Activity propagates forward through the network and is then compared to the correct completion provided by the environment. Note that we do not envision this as an explicit instruction process, but simply a matter of anticipating future inputs from current inputs. It is the mismatch between what the network anticipates and what is provided by the environment that drives the adjustment of connections. In this case, the target pattern specifies the correct set of completions of the item-context pair provided as input. We treat the model as a proxy for the learning children do both from explicit (verbal) propositions provided by others in their environment and from actual experiences with objects in different contexts. For example, the child watching a robin on a branch may not anticipate at first that it will fly away as a cat creeps up upon it. Witnessing that the bird does fly away provides a signal that does not match expectations, and it is the difference between the (null) expectation and the witnessed action that then drives connection-based learning.

The details of the learning process are described in Rogers and McClelland (2004). The process is described as "back-propagation of error," and its biological plausibility has been much maligned, but it has repeatedly been shown how the necessary error information can be derived from temporal differences in activation in networks with bidirectional connections (e.g., O'Reilly, 1996). What is important for our purposes is that all the connections in the network are affected by this process. Some of these connections serve to change how patterns of activation inside the system affect the output units. Others serve to influence how external inputs (on the left in the diagram) are internally represented. The changing structure in these representations is crucial for the patterns of change that we see in development.

Figure 72.4 presents the patterns of activation seen in the "representation" layer of the network in figure 72.3 at three points during learning. Each histogram bar in the left panel of the figure represents the activation of one of the units in the representation layer of the network for a particular item at a particular time. The set of eight such bars for a particular item at a particular time constitutes that item's internal representation. We can see that initially each unit takes a middling activation value for all items, so that the patterns are not well differentiated. The small differences at this stage largely reflect the initial random noise in the connection weights. In contrast, at the last time point (right column of figure 72.4) it is clear that the patterns have become quite differentiated. Hierarchical clustering analysis (not shown) reveals what can also be seen by eye. The network treats the two different types of fish as very similar, but as quite distinct from the two types of birds; all these are very different from all of the plants. Among the plants, the two trees are very similar and so are the two flowers. The trees and flowers are somewhat differentiated from each other, though not as much so as the birds are from the fish.

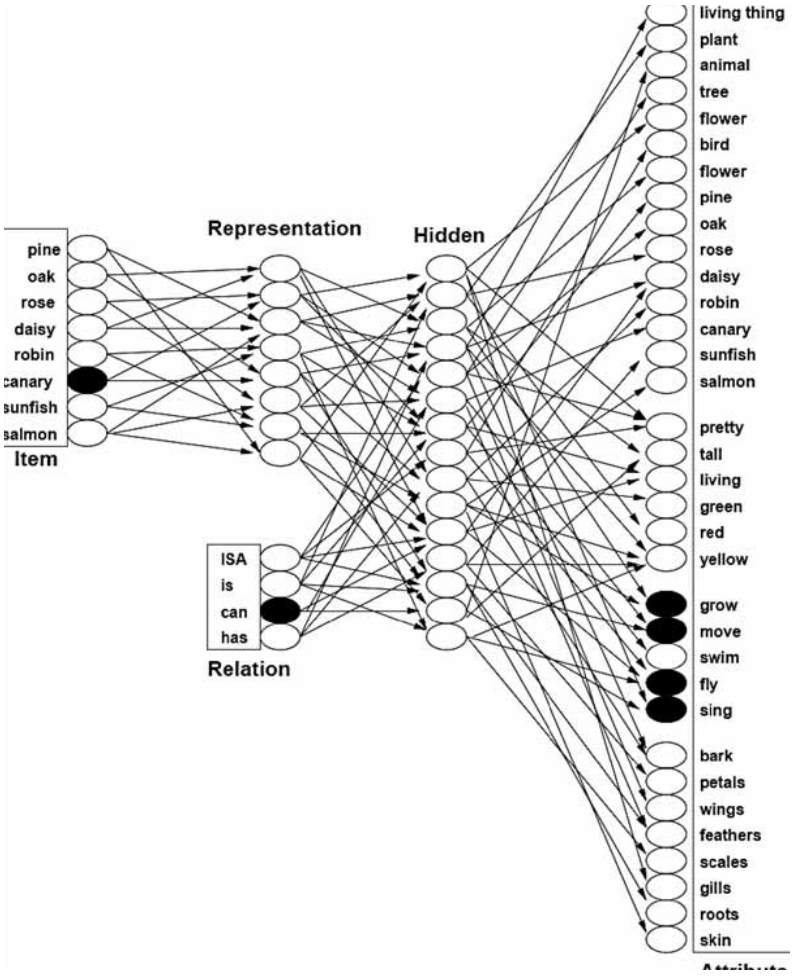


FIGURE 72.3 The connectionist model of semantic memory used in the developmental simulations, adapted from Rumelhart (1990; Rumelhart & Todd, 1993). The entire set of units used in the network is shown. Input units are shown on the left, and activation propagates from the left to the right. Where connections are indicated, every unit in the pool on the left is connected to every unit in the pool to the right. Each unit in the *Item* layer corresponds to an individual item in the environment. Each unit in the *Relation* layer represents contextual constraints on the kind of information to be retrieved. Thus the input pair *canary can* corresponds to a situation in which the network is shown a picture of a canary and asked what it can do. The network is trained to turn on all the units that represent correct completions of the input query and to turn off all other units. In the example shown, the correct units to activate are *grow*, *more*, *fly*, and *sing*. (Reprinted with permission from Rogers & McClelland, 2004, figure 2.2, p. 56.)

What is particularly interesting for our purposes is the fact that the representations undergo a progressive differentiation. Specifically, at the intermediate time point shown, the network has successfully differentiated the plants from the animals before it further differentiates these two superordinate categories into their particular intermediate-level types. The full trajectory of this process is illustrated in a two-dimensonal projection in figure 72.5. An animated version of this figure best captures the progressive differentiation of the network's representations over time. The animation at http://psychology.stanford.edu/~jlm/Presentations/Differentiation.mpg illustrates that the progressive differentiation process is highly stagelike in character, as seen in many aspects of children's cognitive development (McClelland, 1989). In the animation, we can see that all the patterns are initially undifferentiated; they first divide into the plants on the one hand and the animals on the other; then the mammals diverge from the birds; then the trees from the flowers; then the different individuals differentiate from one another.

OVERGENERALIZATION OF NAMES AND ILLUSORY CORRELATIONS The overgeneralization of frequent names and the presence of illusory correlations both arise from the progressive differentiation process. Simulations illustrating the transitory developmental emergence of both these effects are shown in figures 72.6 and 72.7. In figure 72.6 we see what happens in a simulation in which there were four trees, four flowers, four fish, four birds, and five four-legged animals. Among the latter, one—the dog—occurred ten times more frequently than the others, including the goat

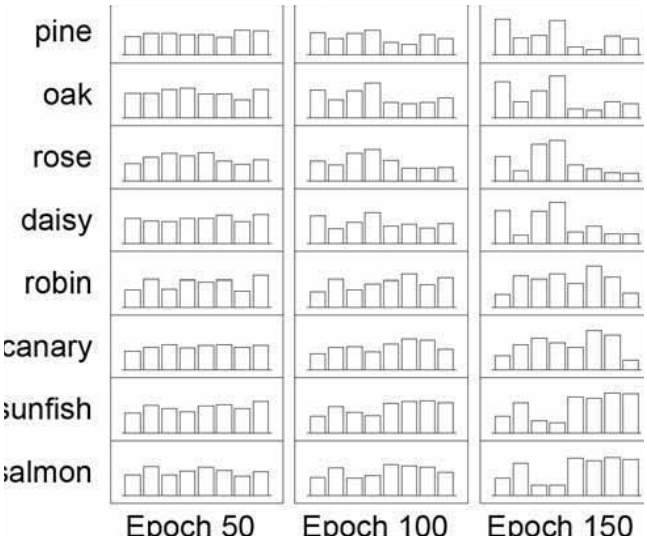


FIGURE 72.4 The process of differentiation of conceptual representations as seen in the Rumelhart model. Learned internal representations of eight items at three points during learning, using the network shown in figure 72.3. The height of each vertical bar indicates the degree of activation for one of the eight units in the network's *Representation* layer, in response to the activation of a single *Item* unit in the model's input. Early in learning (50 Epochs), the pattern of activation across these units is similar for all eight objects. After 100 epochs of training, patterns have begun to differentiate at the superordinate level (plants versus animals) but not at the intermediate level (trees versus flowers, birds versus fish). This further differentiation is apparent after 150 epochs, and continues down to the subordinate level as training continues. (Reprinted with permission from Rogers & McClelland, 2004, figure 3.1, p. 86.)

(consistent with input children receive, Rogers & McClelland, 2004, chapter 5). In this case, the network had a tendency, at a certain point in its development, to activate the name "dog," not only for the dog but also for the other four-legged animals, including the goat. This tendency arises at a time when the four-legged animals are differentiated from the other animals but are not yet well differentiated from each other, and falls away again as the different land animals pull apart. This tendency makes sense from an optimal inference point of view, if we consider the conditional probability of hearing the label "dog" when experiencing the internal representation shared by the land animals (this probability is $10/(10 + 1 + 1 + 1 + 1)$ or about 0.7). Thus when the network has only one shared internal representation for all land animals, its best guess is to call them all "dog." Once the representations differentiate, however, the conditional probability of each label given each item's representation changes dramatically, leading the network no longer to overgeneralize.

In figure 72.7, we see the tendency for the network to activate the property "has leaves" for the pine tree. For the network, this is an illusory correlation; although all other plants have leaves, the pine tree does not in the network's training environment. There is a phase, relatively early in the network's development, where it attributes leaves to all things with a middling activation value (this value represents the proportion of all objects that have leaves). When the plants begin to differentiate from the animals, however, the network attributes leaves more strongly to all the plants, including the pine tree. Only as it comes to differentiate the pine from the other plants does it reverse this illusory correlation. Once again, this pattern makes sense from an optimal inference point of view. At the point in development where the representation of the pine tree is identical to the representation of the other plants, the conditional probability of "leaves" is very high. Again, once the representation of the pine differentiates from the representation of each of the other plants, the conditional probabilities change. Now, the network can learn the conditional probability of leaves for pine is actually 0.

SENSITIVITY TO COHERENT COVARIATION AND ITS DEPENDENCE ON THE ARCHITECTURE OF THE NETWORK In Rogers and McClelland (2004) the reasons why these three phenomena occur in the network are extensively explored. The essential point of this analysis is the observation that the connection weights from the item units to the representation units—and in consequence, the patterns of activation assigned to each concept on the representation layer—are sensitive to the pattern of coherent covariation of properties across the items presented to the network. The fact that all the animals share one set of properties that none of the plants have, while the plants share another set of properties that none of the animals have, is responsible for the first wave of differentiation. The error signals reaching the representation layer (and therefore driving the connection weights) tend to push the representations in a similar direction for all the animals—a direction that is different from the direction in which the error signals push the representations of all the plants. The subsequent differentiation of the different types of animals is also driven by the fact that each type of animal shares a set of properties that none of the other types possess, and similarly for the differentiation of the different types of plants. Now, it is the connections forward from the representation layer to the output layer that determine what output the network generates when a given pattern is present on the representation units. During the phase of development when the network is essentially representing the dog and all the other land animals as the same, but different from the plants and from the other types of animals, the correct name response when this shared representation is present over the representation layer is usually "dog," because the dog occurs more frequently than any of the other land animals. The error-correcting learning process thus pushes the weights forward from this representation layer to activate the name

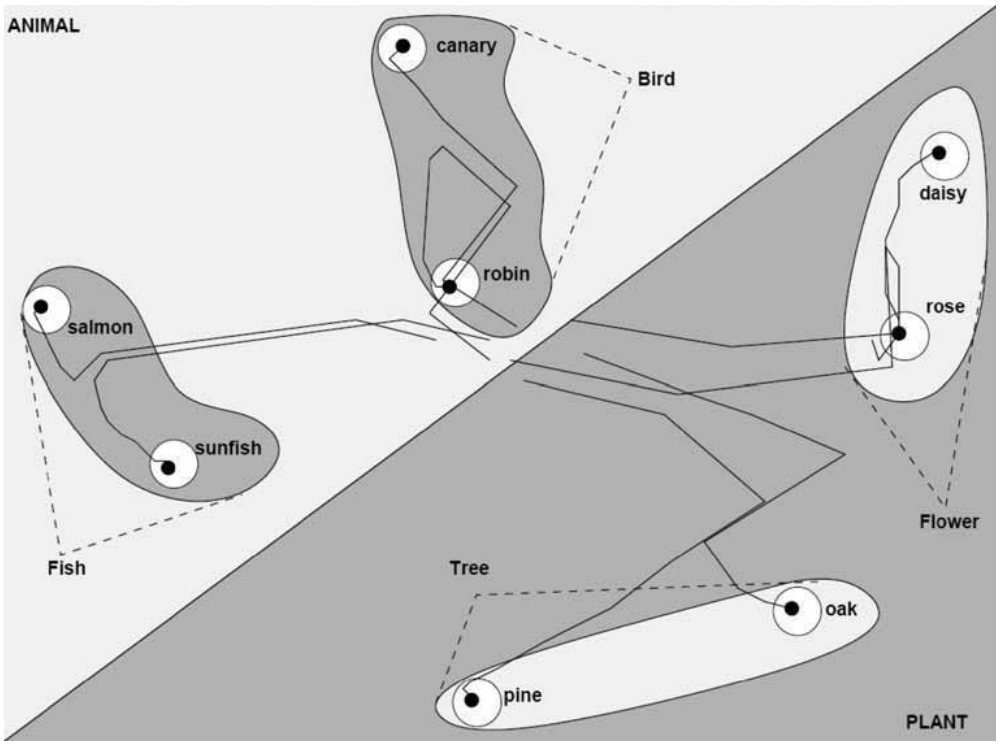


FIGURE 72.5 Trajectories of item representations (lines fanning from center of space) and their ultimate positions in the Rumelhart network's semantic representational space. Shaded regions around the final points indicate the approximate size of regions associated with subordinate, intermediate, and superordinate levels as indicated. When an item has higher frequency, the region associated with that item and its properties is increased. (Reprinted with permission from Rogers & McClelland, 2004, figure 3.9, p. 112.)

"dog" more often than any other name, thereby accounting for the tendency for all land animals to be called "dog" at this stage of development. In the case of attributing leaves to the pine tree, the situation is similar. As long as the representation of the pine tree is similar to the representation of all other plants, the network tends to attribute "has leaves" to it as it does to the other plants. Thus both overgeneralization of frequent names and illusory correlations arise as a consequence of the sensitivity of the representations in the network to coherent covariation.

It is important to see that the network's tendency to be sensitive to coherent covariation, and thus to exhibit differentiation, name overextension, and illusory correlations, is a feature of its architecture (figure 72.8). In the extreme, if each item-context pair projected to its own distinct representation unit, which in turn projected forward to the appropriate properties for the object (as in figure 72.8*B*), there would be no sensitivity to coherent covariation at all. The error signals driving learning of each property of each item would be completely segregated from those relevant to every other property. In our actual architecture (figure 72.8*A*), something very different happens. Because the error signals for each property of each object in each context are projected on the same set of representation units, and because different concepts share these representation units, what is learned about an object in one context tends to be shared both across objects and across contexts. In short, a key observation from our theory is this:

> Sensitivity to coherent covariation, a tendency central to explaining many aspects of conceptual development, requires the use of a shared representation layer mediating all aspects of conceptual knowledge of all different kinds of things.

The particular architecture of the Rumelhart network is only one architecture with this property; in the rest of this article we will be considering networks with slightly different architectures that still make use of a single shared representation mediating all kinds of knowledge of all kinds of things.

Disintegration of conceptual knowledge in semantic dementia

Although the network discussed so far is quite abstract, the simulations do suggest an important hypothesis about the architecture of the cortical semantic network: that there must exist some place in the network where all different kinds of information *converge*, so that different items and events, regardless of the modality of input or the particular semantic domain, get processed through the same set of

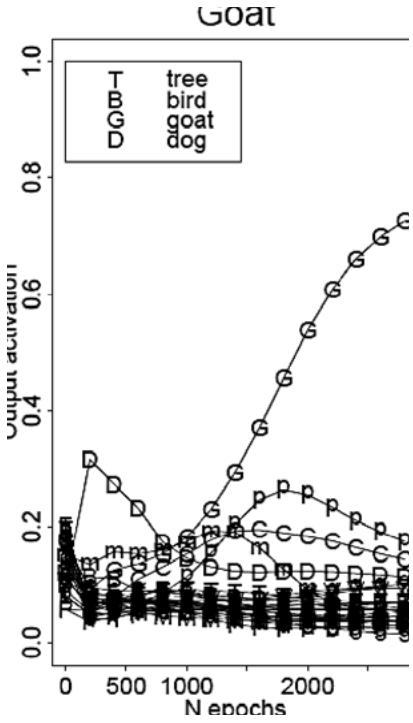


FIGURE 72.6 Activation of all the name units when the model is probed for its knowledge of basic names, in a network trained with dog patterns eight times as frequently as other mammal patterns in the environment. Early in learning, the network tends to inappropriately activate the name "dog," especially for related objects. (Reprinted with permission from Rogers & McClelland, 2004, figure 5.1, p. 214.)

neurons and synapses. It is precisely this convergence of information that promotes sensitivity to coherent covariation in this network. This same convergence can explain the striking pattern of semantic deficits seen in patients with a rare neurological disorder, semantic dementia.

CHARACTERISTICS OF THE DISORDER Semantic dementia (SD) is a progressive deterioration of conceptual knowledge in the context of otherwise greatly preserved cognitive function (Snowden, Goulding, & Neary, 1989; Hodges, Patterson, Oxbury, & Funnell, 1992). Patients show serious deficits in any task requiring them to access knowledge about any type of thing from any form of input (Bozeat, Lambon Ralph, Patterson Garrard, & Hodges, 2000). Thus patients are impaired at both comprehending and producing speech, recognizing words and line drawings of common objects, indicating the correct color for black-and-white drawings of familiar items, choosing which of a set of items makes a particular sound, matching objects on the basis of shared function or use, demonstrating the use of everyday objects, and even recognizing common odors. These impairments affect knowledge for all different kinds of concepts—living and nonliving, abstract and concrete, verbs and nouns—and are apparent regardless of the modality of reception or expression tapped by a particular test. Despite these serious disabilities, patients with SD perform normally or near-normally in tests of basic perception, episodic and working memory, executive function, problem solving, and attention; and, apart from word-finding difficulties arising from their conceptual deficits, they produce fluent and grammatical speech. Such patients thus appear to exhibit a pure and progressive cross-modal and domain-general impairment of semantic or conceptual knowledge (Patterson, Nestor, & Rogers, 2007; Lambon Ralph & Patterson, 2008).

As notable as the syndrome itself is the remarkable anatomical specificity of the cortical atrophy observed in the disease, which without exception affects anterolateral regions of the temporal lobes. Bilateral degeneration is the norm, though pathology is usually asymmetrical and patients with left-predominant atrophy present at about double the rate of right-predominant cases. The typical pattern on structural MRI is well-defined atrophy of both anterior temporal lobes that is maximal at the temporal pole and on the adjacent rostral-inferior surface (Hodges & Patterson, 2007; Patterson et al., 2007).

The pattern of semantic dysfunction together with the anatomical specificity of the atrophy provide strong evidence that the cortical semantic network adheres to the "convergence principle" suggested by the modeling work discussed earlier. Specifically, the findings from SD suggest that semantic knowledge for all kinds of concepts, across all modalities of reception and expression, depends upon a relatively circumscribed region of the anterior temporal lobes. Perhaps, then, these regions play a functional role similar to that of the *representation* layer of the Rumelhart network: by processing information about all kinds of items in many different situations and contexts, perhaps these regions form learned internal representations of inputs that capture the semantic similarity among items.

MODELING SEMANTIC DEMENTIA Rogers, Lambon Ralph, Garrard, and colleagues (2004) investigated this hypothesis using a PDP model based on the architecture shown in figure 72.9. Here, different kinds of sensory, motor, and linguistic information are represented in different pools of units, with each pool dedicated to a particular kind of information. These *surface* representations receive direct input from the corresponding sensory systems so that, whenever a given stimulus is encountered, the units that code its directly observed properties are activated. Like many other researchers, we believe the different surface representations to be subserved by different brain areas, organized predominantly by modality, and situated near the sensory channels from which they receive input (Martin & Chao, 2001). Also in common with other researchers, the function of the semantic system in this framework is to mediate learned associations among the various different surface

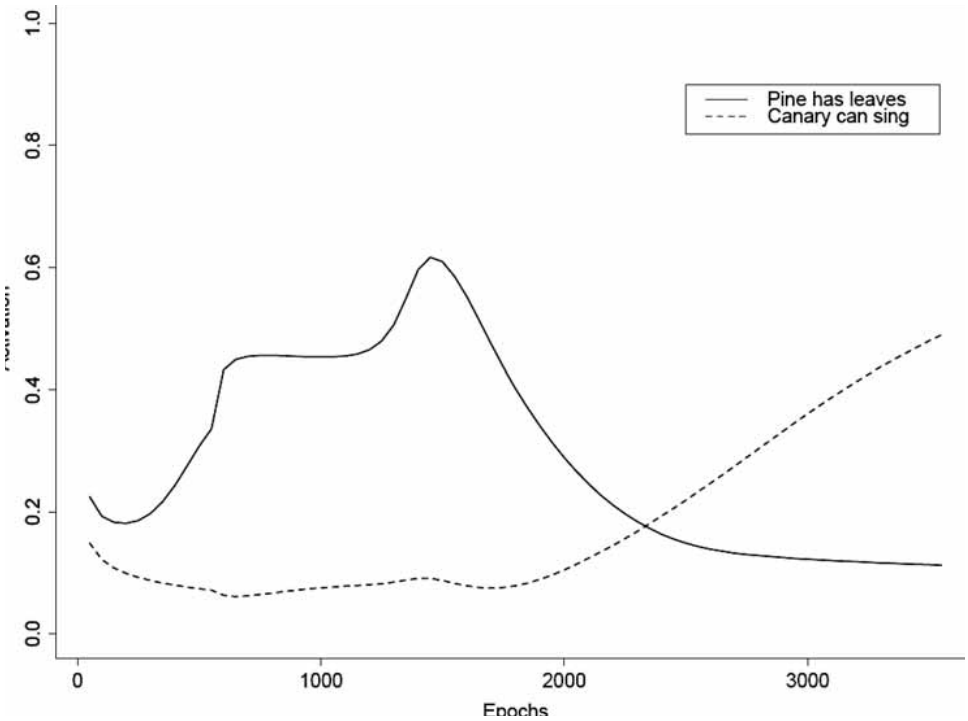


FIGURE 72.7 The activation of the "has leaves" and "can sing" output units across the first 5000 epochs of training, when the network is probed with the inputs *pine has* and *canary can*, respectively. At epoch 1500, the network has been trained 150 times to turn off the "has leaves" unit in response to the input *pine has*, and to turn on the unit "can sing" in response to the input *canary can*. Nevertheless, the network still activates the "has leaves" unit for the pine tree, and fails to activate the "can sing" unit for the canary. (Reprinted with permission from Rogers & McClelland, 2004, figure 6.6, p. 254.)

representations—so that, for instance, when a line drawing of a pencil is observed, representations of associated attributes in other modalities, such as the appropriate color (yellow), name ("pencil"), and action (writing), will also become activated in the appropriate modality-specific brain areas.

Our framework differs from some others in proposing that the associations between all different forms of surface representations are mediated by a central "hub" in the anterior temporal lobes. The hub itself receives no direct input from sensory systems, but receives connections from and sends connections to the surface representations that code representations of particular sensory, motor, and linguistic attributes. The hub, illustrated in figure 72.9, is similar to the *item representation* layer in the Rumelhart network in that the representations there are a consequence of the learning process that shapes the weights projecting into and out from the layer. The surface representations in figure 72.9 are analogous to the *attribute* output units in the Rumelhart model, in that they explicitly encode observable characteristics of items in the environment. Because processing is recurrent in this model—activation flows both from the surface representations to the hub and from the hub back to the surface—there is no need to have separate "input" and "output" layers. Instead, any given input can be specified as a distributed pattern of activation across corresponding units in the surface layers of the model. These inputs will propagate activation up to the hub of the anterior temporal lobes, which will then feed activation back to the surface representations to activate other properties of the item that have not been directly observed.

This framework also provides a natural paradigm for learning. The top-down activation of surface attributes may be viewed as the generation of an implicit expectation about the item's unobserved properties. If these expectations are contradicted by a subsequent observation, the discrepancy can be used as an error signal to drive weight changes throughout the network, so that the system comes to generate increasingly accurate expectations.

Rogers, Lambon Ralph, Garrard, and colleagues (2004) used a model implementing these ideas to assess whether the theory could account for patterns of semantic impairment observed in SD. The model consisted of a *Visual* layer in which each unit coded a visually apparent property of an object; a *Verbal* layer, in which each unit coded a predicate (e.g., "has wings") that might appear in a verbal description of an object, including names and other descriptors; and a *Semantic* layer that mediated interactions between these. Visual perception of an object was simulated by directly activating the object's properties in the Visual layer and allowing activation to propagate through Semantic units to

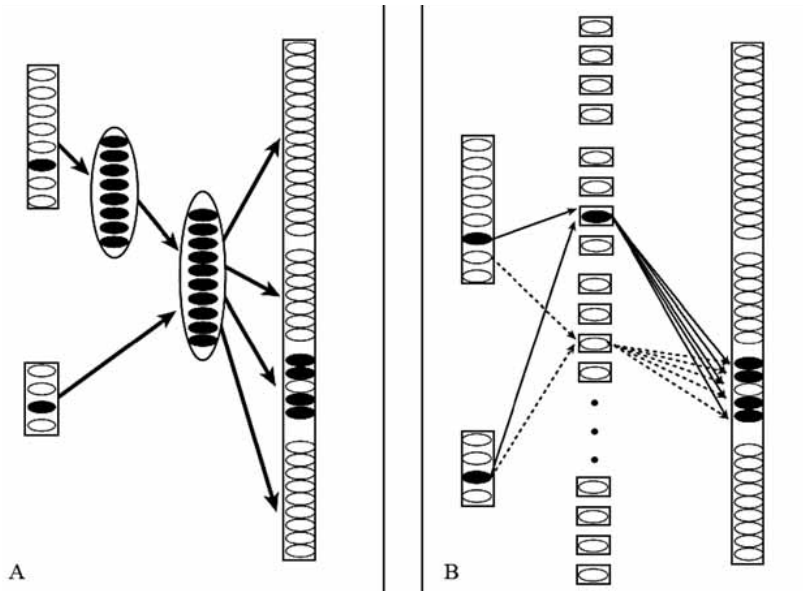


FIGURE 72.8 Two alternative feedforward architectures for mapping from localist *Item* and *Context* inputs to sets of output properties. Thick arrows in *A* indicate full connectivity among units in the sending layer to those in the receiving layer; thin arrows in *B* indicate individual connections. The shading of units indicates how activation spreads forward given the input *canary can*. (*A*) The Rumelhart network architecture, in which all items first map to a single context-independent hidden layer and then converge with context inputs in a second hidden layer. The first layer of hidden units receives error signals that are filtered through the second, convergent representation; hence the architecture constrains the network to find context-independent representations of individual items that are sensitive to coherent covariation of properties across different contexts. (*B*) Separate localist representations for every possible conjunction of item and context. Only connections for *canary can* (solid arrows) and *salmon can* (dotted arrows) are shown. In this case the model will not generalize and will not be sensitive to coherent covariation. (Adapted with permission from Rogers & McClelland, 2004, figure 9.1, p. 357.)

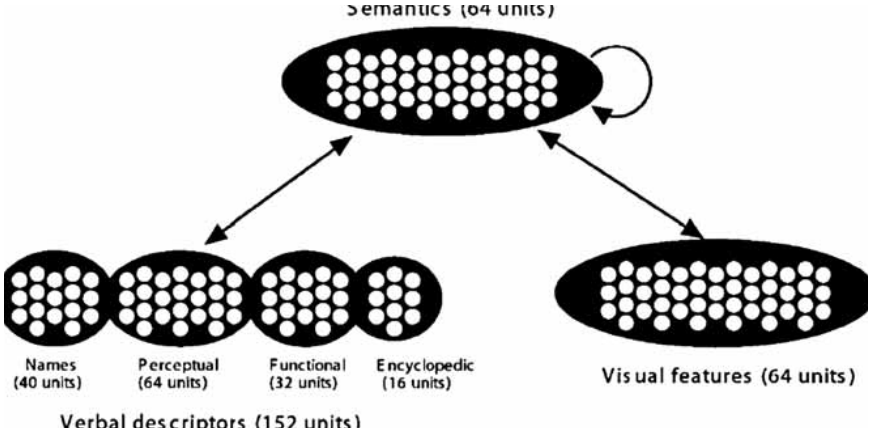


FIGURE 72.9 Architecture of the model used to simulate semantic dementia. (Reprinted with permission from Rogers, Lambon Ralph, Garrard et al., 2004, figure 1, p. 207.)

Verbal units. Presentation of an object name was simulated by directly activating the single unit representing the name in the Verbal layer; and presentation of a verbal description of an object was simulated by activating the subset of units representing the presented predicates. The model was then trained with backpropagation to produce the correct visual pattern, verbal pattern, or name when provided with one of these representations as input. Representations on the Semantic layer were not specified but, as in the Rumelhart network, emerged as a consequence of learning. Once trained, the model permitted simulation of the most basic tasks used to assess semantic memory in patients with SD, including visual object naming, drawing-to-name, delayed-copy drawing, word-to-picture matching, sorting words and pictures, and so on. The patterns used to train the model were constructed to capture important aspects of similarity

structure apparent in both verbal attribute-listing studies and drawings of common objects. In both cases, superordinate category structure was clearly apparent—items from the same superordinate category (e.g., animals, manmade objects, and plants) tended to share many verbal descriptors and also to have similar visible parts in their drawings. There was also some more specific structure, especially among the set of animals; for example, different birds were more similar to each other than they were to other types of animals.

After training, the model represented each individual item—whether accessed by means of a single name, a verbal description, or a visual pattern—with its own domain-general pattern of activation across Semantic units. Just as with the Rumelhart network, these patterns captured similarity relations, with semantically related items represented by similar patterns of activation.

The deficit in semantic dementia was simulated by removing an increasing proportion of the connection weights projecting into or out from the intermediating hidden layer; the model was then tested on analogs of the semantic tasks used with patients. Comparable results are obtained by deleting individual units rather than individual connections.

The model naturally captures both the cross-modal and cross-category nature of the semantic impairment in SD. Semantic task performance was compromised both for visual semantic tasks like delayed-copy drawing—where the participant must draw a previously viewed item from memory after a short delay—and for cross-modal tasks like object naming; and the magnitude of impairment was roughly equivalent for animals and manmade objects (Lambon Ralph, Lowe, & Rogers, 2007). These aspects of model performance follow from the fact that it respects the convergence constraint identified in the previous section. When these weights were lesioned, performance for all tasks and semantic domains was affected.

Parallels Between Development and Disintegration

The model also offers a clear explanation of what for us is one of the key features of semantic dementia: loss of differentiating details about particular concepts together with spared knowledge of more general information. This loss of differentiating detail was first documented in the initial report of semantic dementia by Warrington (1975). Warrington also pointed out the parallelism between this finding and the lack of differentiation in early stages of conceptual development. Warrington demonstrated that knowledge of an item's properties that characterize broad semantic categories—for instance, the fact that tigers have fur—is much more robust than knowledge of item-specific properties, such as the fact that tigers have stripes. Knowledge about properties that characterize very specific classes—for instance, particular breeds of dog—is much more vulnerable to early impairment, as is knowledge about less frequent and less prototypical items. So, when categorizing familiar objects, patients with very mild impairment will usually fail at naming objects at subordinate levels such as "robin" or "BMW," but can seem unimpaired at more general levels such as "bird" or "car"; and even very semantically impaired individuals can succeed as well as controls at very general categories such as "animal" or "vehicle" (Rogers & Patterson, 2007).

In the model described by Rogers, Lambon Ralph, Garrard, and colleagues (2004), this erosion of knowledge about the individuating details of specific concepts arises as a consequence of the similarity structure of the distributed representation acquired in the semantic layer. To see this point, consider how the healthy model retrieves a fact specific to a particular subordinate concept, such as that a robin has a red breast. Even though Rogers and colleagues used the architecture shown in figure 72.9, the situation is still well reflected in figure 72.5, which shows the situation arising in the simpler training environment used in the Rumelhart network. What we see in this figure is a graphical depiction of the fact that the models learn to treat the robin as quite similar to all the other birds (in this case, the only other bird is the canary); yet none of these other birds has a red breast like the robin. Therefore, to correctly activate the "red breast" units in either the Visual or Verbal layer, the Hidden layer must instantiate the pattern of activation corresponding to the robin almost exactly—if this pattern is just a little different, then it may become more similar to another individual bird that does not have a red breast, and the system will fail to activate the correct property in the periphery. Thus relatively small distortions to the correct representation will prevent the model from strongly activating properties that are unique to very specific categories.

Now consider a property that the robin shares with other birds, like "has wings." In this case, the model need not instantiate precisely the right representation to retrieve the property, since it is common to all of the birds and hence will be activated by all of the patterns that are somewhat similar to the correct pattern. If the robin representation is distorted so that it more closely resembles the canary representation, this distortion will not disrupt activation of the "has wings" units in the Visual and Verbal layers, because the canary has wings just like the robin (as do all of the other birds). Thus even with a relatively severe distortion to the representation the system can still generate the correct outputs for category-typical properties. The same argument suggests why still more general properties—like the fact that animals have eyes—are even more robust. If the robin representation is, as a consequence of brain damage, so degraded that it becomes less distinguishable from the various mammals as well as from the other birds, the system will still continue to correctly activate properties held in common between birds and mammals.

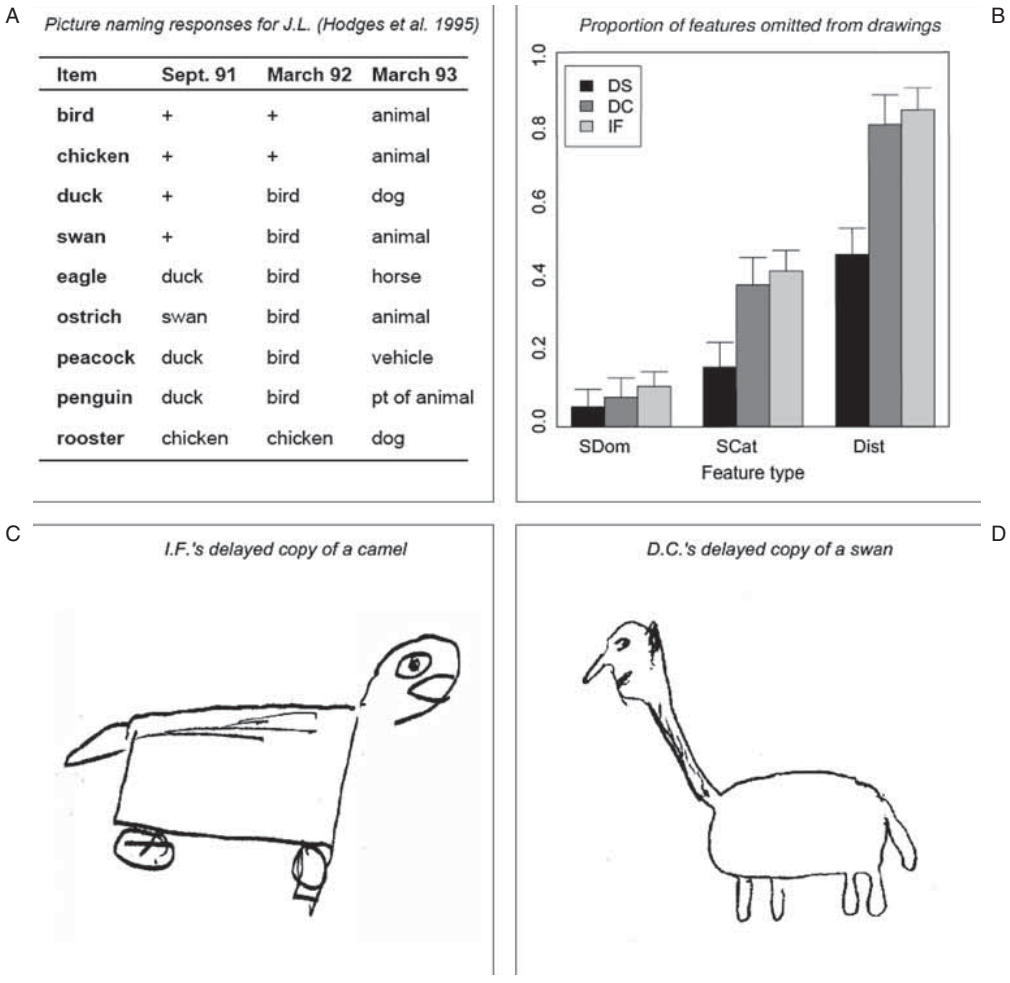


FIGURE 72.10 Evidence of conceptual disintegration in semantic dementia. (*A*) Naming responses given by patient JL to pictures of birds (drawn from a set of line drawings for which control subjects consistently provide the name given in the left column) at three times during the progression of his illness (+ indicates correct responses). (*B*) Proportion of features of different types omitted from drawings by three other semantic dementia patients. Patients were shown a picture of the object including all the tested properties and were asked to copy the picture from memory after a 10-second delay. All copied the picture accurately while it remained in view, but had difficulty in reproducing the distinctive but not the domain-general properties of the pictured objects after a delay. SDom: Properties shared by typical members of the general domain (e.g., eyes, shared by animals) of the test item. SCat: Properties shared by typical members of the superordinate category (e.g., wings, shared by birds). Dist: Distinctive features of the test item itself (e.g., stripes, distinctive attribute of tiger). (*C*) Delayed copy of a camel; no hump is evident. (*D*) Delayed copy of a swan. A long neck is present, indicating some preserved representation of specific information but there are four legs, illustrating the tendency these patients have to fill in properties that are generally present in items within the overall domain (animals) even if not present in the specific item (swan) or even its immediate superordinate (bird). (Redrawn with permission from McClelland & Rogers, 2003, figure 2, p. 312. Panel *A* excerpted from the appendix of Hodges, Graham, & Patterson, 1995, p. 490.)

The gradual loss of idiosyncratic detail coupled with the preservation of information shared among members of a category creates progressive dedifferentiation of semantic knowledge, paralleling the progressive differentiation seen in development. Hand in hand with this progressive dedifferentiation are two other phenomena that parallel those seen in development: (1) overgeneralization of names of frequently occurring objects and (2) illusory correlations, or the attribution of category- or domain-general properties to objects that lack these properties. Evidence of these aspects of SD is shown in figure 72.10. In figure 72.10*A*, we present picture-naming data from patient JL at different stages of his progressive deterioration (Hodges, Graham, & Patterson, 1995). Here we see that, as his impairment becomes progressively worse, he shows an increasing tendency to overapply names of the more common animals (e.g., duck) to less common animals (eagle, peacock). In figure 72.10*C* and 72.10*D*, we present delayed copies made by two patients of a swan and a camel. In the latter case, the differentiating detail of the camel's hump is lost; more strikingly, in the former, a property typical of the broad class of animals—that of having four legs—is added to the swan, making it far

more like other animals. These phenomena are not idiosyncratic to particular objects or patients (figure 72.10*B*). There is a general tendency to produce names of more frequent category coordinates in object naming (Woolams, Cooper-Pye, Hodges, & Patterson, 2008) and a general tendency to both omit differentiating details and to incorrectly incorporate domain-general properties in patient's delayed copying (Bozeat et al., 2003).

The reasons for overgeneralization of frequent names and illusory correlations can again be understood with reference to figure 72.5. As illustrated in the figure, the points in representational space associated with specific items are a function of the item's frequency. A very small region surrounding a very specific point is associated with the idiosyncratic properties of individual objects, including their names, whereas a much larger region is associated with the properties of frequent objects (such as a dog) or the shared properties of many objects (such as having four legs). As a result, any distortion of the representation of a particular, relatively uncommon item will tend to result in the network's representation landing in a part of the space associated with more common objects and more typical properties. Indeed, the model shown in figure 72.9 was able to simulate closely the proportions of category coordinate naming errors seen in SD, as well as the proportion of item-specific omissions and category-general overgeneralization errors made by such patients (Rogers, Lambon Ralph, Garrard, et al., 2004).

In summary, just as the Rumelhart model explains the progressive differentiation of conceptual knowledge over development, the model of Rogers, Lambon Ralph, Garrard, and colleagues (2004) explains the apparent reversion of this process in SD: the gradual erosion of knowledge about the details that individuate concepts, beginning with very specific concepts and progressing to more and more general concepts. All these phenomena arise from the same general principles: that semantic knowledge is acquired through domain-general mechanisms, in a system that learns mappings among various different "kinds" of sensory, motor, and linguistic information and that stores these mappings within a convergent architecture in which all kinds of information for all kinds of concepts are processed through the same neurons and synapses.

Nonsemantic deficits in semantic dementia

The previous section demonstrates how the essential principle of convergence from widely distributed modality-specific brain regions can help us to understand the *semantic* deficits observed in semantic dementia (SD). This section addresses a *further* set of deficits observed in SD. These deficits are important because the abilities affected are ones that have been thought by many to operate *without* reference to semantic knowledge. If SD constitutes a semantic impairment, why should these abilities be affected? One possibility is that the nonsemantic impairments are simply additional independent deficits that arise because of abnormalities in other, nonsemantic regions. However, these additional deficits both (1) occur consistently with the semantic deficit and (2) are similar in nature to the semantic deficit. From these observations, we have argued that they are a part of the core deficit itself (e.g., Patterson et al., 2006). We concentrate here on "nonsemantic" SD deficits in four tasks using words, though similar deficits occur with other kinds of stimuli. Nearly all SD patients had abnormal performance on each of the four tests: (1) lexical decision, in which the participant must judge whether each of a series of letter strings is a real word or must choose the real word when items are presented in pairs; (2) oral reading of single printed words; (3) written spelling of single spoken words; and (4) oral production of the past tense from present tense (stem) forms of verbs (Benedet, Patterson, Gomez-Pastor, & de la Rocha, 2006; Funnell, 1996; Graham, Patterson, & Hodges, 2000; Hodges et al., 1995; Patterson & Hodges, 1992; Patterson, Lambon Ralph, Hodges, & McClelland, 2001; Rogers, Lambon Ralph, Hodges, & Patterson, 2004; Ward, Stott, & Parkin, 2000). The basis for performing three of these tasks (reading, spelling, and past tense formation) is traditionally considered to be a joint function of a system of rules and a system of lexical entries; the rules and lexical entries are often considered separate from each other and also separate from semantic knowledge (e.g., Caramazza, 1997; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Levelt, 1989; Pinker, 1991). Lexical decision is often thought to depend only on lexical entries, although the presence of typicality effects in this task might be taken to suggest that some other system of knowledge, perhaps a system of rules, would be needed here, too.

Within the PDP framework, performance on all these tasks is thought to rely on a single integrated processing system that contains neither rules nor lexical entries but, like the semantic system, is sensitive both to properties that items share with each other and to idiosyncractic, item-specific information (Rumelhart & McCllelland, 1986; Plaut, McClelland, Seidenberg, & Patterson, 1996). Words, like objects, tend to have properties that they share with others. Tigers have fur like other animals, but they also have their idiosyncratic stripes. Similarly, the word *pint* has correspondences it shares with many other words in the pronunciation of most of its letters, but it has an idiosyncractic correspondence in the pronunciation of the vowel; and the irregular verb *keep* forms its past tense like regular words in most respects. In regular verbs, a /t/ would be added to the stem (cf. *bake–baked*, pronounced /be:kt/). The same is true with *keep–kept*, but in addition there is an idiosyncratic vowel adjustment.

Semantic dementia patients have difficulty in all of these tasks in ways that parallel the deficits they show in semantic tasks. As their semantic disorder progresses, they make progressively more errors on words that are of low typicality, especially when they are also of low frequency. Also, the nature of the patients' errors mirrors precisely the "illusory correlations" discussed earlier with reference to knowledge of the properties of objects. Semantic dementia patients apply to atypical words the correspondences that the words would have if they were more typical. Asked to read the word *sew*, they often say "sue" using the typical correspondence in *crew, few, new*, and so on; asked to spell "cough," they write *coff*, using the typical spelling for /f/ as in *off, scoff, fluff, cuff*, and so on. Asked to put the sentence "Every day I fight with my brother" into the past tense, they will often say, "Yesterday I fighted with my brother." A typicality effect is also seen in lexical decision. Asked which of the two letter strings *seize* and *seese* is a real word, severe SD patients actually prefer the incorrect but more typical spelling *seese*. The effect is strictly analogous to an effect seen in an object-decision task (Rogers, Lambon Ralph, Hodges, & Patterson, 2004). Asked to choose between a real elephant with large floppy ears of the kind that one only sees on elephants and an otherwise identical elephant with smaller, more typical ears taken from a monkey, severe SD patients will tend to choose the pseudoelephant over the real one.

A Single System for Knowledge of Both Objects and Words Clearly, semantic dementia patients show a conjunction of semantic and lexical deficits. These deficits not only correlate (Graham et al., 2000; Patterson et al., 2006; Woollams, Lambon Ralph, Plaut, & Patterson, 2007) but also exhibit similar characteristics. In semantic tasks, the patients tend to lose knowledge of specific and idiosyncratic properties of objects while retaining and overextending knowledge of general and typical properties. Similarly, in lexical tasks such as word reading, word spelling, and lexical decision, they tend to lose knowledge of atypical items and mappings while retaining and overextending knowledge of the typical (as regularization errors). The strong correlation of these deficits and the similarity of their nature suggest that the tasks all depend on the same set of processing structures that also underlie semantic processing. On this view, it is damage to these structures that produces both the semantic deficits and the parallel deficits in lexical tasks seen in SD patients.

Dilkina, McClelland, and Plaut (2008) developed a connectionist model implementing just such a single-system approach to semantic and lexical processing and showed that it could reproduce the convergent pattern of semantic and lexical deficits seen in SD. The simulation focused primarily on one "semantic" task—picture naming—and one "lexical" task, word reading. The model builds on the one used by Rogers, Lambon Ralph, Garrard, and colleagues (2004). In place of individual units to represent printed or spoken words, Dilkina and colleagues used patterns of activation over orthographic (letter) and phonological (sound) units for the spellings and sounds of words, respectively (figure 72.11). Like Rogers and colleagues, the model uses a single cross-modal level of representation that integrates all types of information about both words and objects and thus corresponds to the convergent semantic representations of the

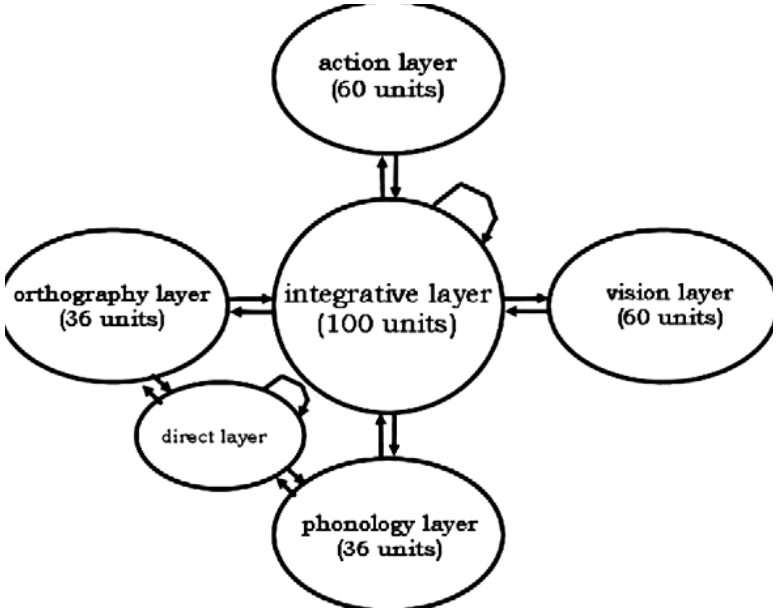


Figure 72.11 Network architecture used to simulate a single integrative system for semantic and lexical processing. (Reprinted with permission from Dilkina, McClelland, & Plaut, 2008, figure 3, p. 143.)

earlier model. The model also builds on earlier work (Plaut et al., 1996; Seidenberg & McClelland, 1989) by including a "direct" route between spelling and sound, in addition to bidirectional connections between both spelling and sound and the semantic layer. Although the direct route tends to specialize in capturing typical spelling-to-sound correspondences whereas the pathway through the integrative layer tends to specialize in idiosyncratic word-specific information, the partitioning is not absolute, and neither pathway corresponds to a strict rule system or a strictly lexical system.

The model was trained to map among four surface representations: (1) visual representations of what entities (objects and animals) look like; (2) action representations of how one may interact with these entities; (3) phonological representations of their names; and (4) orthographic representations of these names. The visual and action representations were binary vectors based on probabilistic category prototypes similar to those used in Rogers, Lambon Ralph, Garrard, and colleagues (2004). The name representations were simple onset-vowel-coda patterns that approximate English spelling-sound consistencies. After the network was trained to map from either visual or orthographic input to produce all four surface patterns, the semantic layer was progressively damaged to simulate semantic dementia.

The model exhibited the overall characteristic performance of SD patients—strong frequency effects in both tasks, and a frequency-by-typicality interaction in reading, as well as a high correlation between naming and reading of irregular words. In addition, the model was applied to the specific pattern of deficit seen in five patients tested with the same set of materials, including the case of an SD patient (patient EM; Blazely, Coltheart, & Casey, 2005) who showed spared reading of low-frequency exception words in spite of a fairly profound impairment in standard semantic tasks. The model was able to fit the specific pattern of reading and naming data observed in all five patients, including patient EM.

Patients like EM and others with similar deficits (Cipolotti & Warrington, 1995) have been used to argue against the single-system account, and thus it is important to understand how the model was able to address this patient, while at the same time addressing the other four, more typical, cases. The basis for this approach lies in incorporating the assumption that there are both premorbid and postmorbid *individual differences* that can contribute to the detailed pattern of performance seen in individual SD patients. Dilkina and colleagues (2008) focused on three such factors—premorbid experience with reading, premorbid capacity of the neural substrate mapping visual word form to phonological word form, and the spatial distribution of the lesion. Each of these factors was motivated by previous literature strongly suggesting that people indeed vary along these dimensions. The three factors were independently manipulated in the model. Each of them significantly and independently affected the relative robustness of naming and reading. Notably, the model was able to successfully fit the SD dissociation case EM by manipulations that made reading relatively more robust to damage than naming. Notably, an experience manipulation alone was sufficient. Even though the model posits that naming and reading involve a single underlying system, greater premorbid experience with reading makes this system's reading performance more robust under damage. In such cases, while naming declines quickly, the decline in reading may be delayed. Similar effects occur with other manipulations that increase reading robustness or that distribute the lesion toward connections from visual more than orthographic input. The account predicts that, as patients like EM progress in their illness, the deficit will eventually affect reading in all cases. Where data are available to test this prediction, it has held up to date (Woolams et al., 2007).

In summary, Dilkina and colleagues (2008) provide a theoretical and a computational account of how semantic and lexical deficits arise within a single system, and how they may appear to dissociate in some cases. While this work does not rule out a separate-system account, it shows that it is not necessary to postulate two systems to explain a handful of dissociation cases. Moreover, a single-system approach seems much more suitable in light of the large body of evidence showing a highly consistent SD profile where semantic and lexical abilities decline together, where there is a distinct frequency-by-typicality interaction in both domains, and where compromised performance results in homologous types of errors.

Roles of the ATL and other brain areas in semantic cognition

As noted throughout this chapter, our approach to understanding the nature and neural basis of conceptual knowledge has deliberately spanned a wide range of approaches and sources of data. In the preceding sections we reviewed a selection of these, including our overarching theoretical framework for considering conceptual knowledge and how this has been implemented in various PDP models addressing two core sources of empirical data—namely, the development of concepts in children and the structured degradation of concepts observed in semantic dementia. In this final section we review other sources of evidence bearing on our hypothesis that the anterior temporal lobes (ATLs) provide a hub over which conceptual knowledge is represented. First we review convergent evidence for the notion that the ATL is a critical part of the wider brain network that supports semantic cognition. Then we broaden the theoretical canvas to include the role of other brain regions. This allows us to consider a wider range of brain mechanisms that contribute

to semantic cognition, construed broadly as the task-modulated use of semantic knowledge to guide behavior.

Convergent Evidence for the Role of the ATL in Semantic Memory There is considerable debate about the putative role of different brain regions in tasks requiring use of semantic knowledge (Hickok & Poeppel, 2007; Martin, 2007; Patterson et al., 2007; Wise, 2003). As noted earlier, in SD a selective semantic impairment is paired with relatively circumscribed atrophy of the anterior, inferolateral temporal lobes, bilaterally. Thus, as already articulated, the simplest and most obvious hypothesis is that the ATL areas are critical for semantic memory (Lambon Ralph & Patterson, 2008; Patterson et al., 2007; Rogers, Lambon Ralph, Garrard, et al., 2004). Given that SD is a neurodegenerative condition, there is no absolute boundary to the damage, and there is always the possibility that subthreshold damage or dysfunction due to invading pathology occurs elsewhere and that it is this more subtle, widespread damage that is the root of the patients' semantic impairment (Martin, 2007). It is critically important, therefore, to derive convergent evidence about the putative role of ATL regions in conceptual knowledge.

Converging evidence comes in three forms: other patient groups, functional neuroimaging, and transcranial magnetic stimulation (TMS). Other neurological conditions do produce semantic impairment when damage affects the same bilateral temporal lobe regions as semantic dementia. These include Alzheimer's disease (Hodges, Salmon, & Butters, 1992) and herpes simplex virus encephalitis (Lambon Ralph et al., 2007; Noppeney et al., 2007), although the more widespread brain damage associated with these neurological diseases leads to additional cognitive and memory impairments (Lambon Ralph & Patterson, 2008). The functional neuroimaging literature provides a slightly complex picture. If one primarily looks at the fMRI literature, there is a distinct lack of evidence for our hypothesis: fMRI studies of semantic memory or comprehension rarely find activation in anterior temporal lobe regions (Devlin et al., 2000; Garavan, Ross, Li, & Stein, 2000). While there may be important task design issues in some of these studies, the failure to find anterior temporal lobe activation reflects, at least in part, fMRI signal loss and distortion that is particularly pronounced in orbitofrontal cortex and the inferior and polar aspects of the temporal lobes (Devlin et al.; Wise, 2003). Functional neuroimaging that utilizes PET does detect semantically related activation in the anterior temporal lobes, even when the same experiment conducted in fMRI does not (Devlin et al.). Likewise, semantic-related processing in the ATL has been observed in normal participants by using MEG, irrespective of whether the stimulus is presented in the auditory or visual modality (Marinkovic et al., 2003), matching findings from early PET-based studies based on pictorial or verbal input (Vandenberghe, Price, Wise, Josephs, & Frackowiak, 1996).

Of course, the areas differentially activated in imaging studies do not imply a *necessary* role (Price & Friston, 2002). Given the potential doubt over the neuropsychological data, we have recently initiated a new line of investigation that uses off-line, repetitive transcranial magnetic stimulation (rTMS) to probe the role of ATL in neurologically intact participants (Lambon Ralph, Pobric, & Jefferies, 2009; Pobric, Jefferies, & Lambon Ralph, 2007). By using timed versions of the semantic tasks used in SD studies, we have been able to compare the pattern observed in the patients with that seen in normal participants post rTMS. The results closely mirror the characteristics of semantic dementia. For example, ATL rTMS produces a temporary slowing of responses on semantically related tasks (e.g., synonym judgment) but not other cognitive tasks matched for overall difficulty (e.g., number judgment). The same stimulation also affects expressive tasks (picture naming but not number reading is slowed). Intriguingly, like the SD patients, a greater effect was observed for identifying concepts at a specific level (e.g., *golden retriever*) than at a basic level (e.g., *dog*). The relative role of the left versus the right ATL can also be probed using this rTMS method. In one study we compared left versus right ATL rTMS on the same synonym judgment task. A comparable slowing of semantic decision times was observed, indicating that both left and right ATL support semantic memory (Lambon Ralph et al., 2009). This pattern was replicated in a further study on a test of semantic association (e.g., between pyramids and palm trees; Bozeat et al., 2000; Howard & Patterson, 1992). In these assessments, stimuli are presented either as pictures or as written words. Repetitive TMS to either left or right ATL produced equivalent slowing on both the verbal and nonverbal versions of the task (Pobric, Jefferies, & Lambon Ralph, submitted).

These convergent results are all in keeping with our hypothesis that the ATL lobes jointly provide an amodal hub for semantic knowledge. However, there are still some puzzling data. Perhaps the most striking results come from patients with ATL resection for intractable epilepsy, who are not clinically associated with a postsurgical semantic impairment, at least not to the same degree as SD patients (Hermann, Davies, Foley, & Bell, 1999). Future studies of semantic memory that directly compare SD and TLE-resection patients are required to understand if the TLE data are truly inconsistent with our hypothesis. Most of the literature on the sequelae of temporal lobe resection is focused upon episodic memory impairment and anomia (which might itself reflect subtle semantic impairment; Lambon Ralph, McClelland, Patterson, Galton, & Hodges, 2001), and semantic memory is rarely formally tested (Giovagnoli, Erbetta, Villani, & Avanzini, 2005). Where semantic perfor-

mance has been assessed, studies have found subtle multimodal impairments both in unoperated TLE patients (Giovagnoli et al., 2005) and in patients after temporal lobe resection (Wilkins & Moscovitch, 1978). Furthermore, temporal lobe resection is a unilateral procedure, but SD patients have bilateral temporal lobe atrophy; it may be that bilateral damage is required to produce significant semantic impairment. It must also be noted that localization of function is complicated in these patients because long-standing epilepsy might lead to changes in neural organization. Indeed, recent imaging studies have shown that white matter connectivity and neurotransmitter function are significantly altered in this condition (Hammers et al., 2003; Powell et al., 2007). In addition, there might be some postsurgical reorganization that is less likely in neurodegenerative conditions when the brain is subjected to constant brain injury (Welbourne & Lambon Ralph, 2005). Consistent with this hypothesis, Wilkins and Moscovitch (1978) found a negative correlation between the severity of semantic impairment and time postsurgery.

Beyond the ATL: The Role of Other Brain Regions in Semantic Cognition As articulated previously, our theoretical framework proposes one hypothesis about the way in which semantic knowledge is acquired through development and how it breaks down to produce the multimodal semantic impairments observed in semantic dementia and other ATL-focused neurological diseases. However, full-fledged semantic cognition—defined here as the adequate use of semantic knowledge to guide complex behavior—requires not only the ability to activate stored information from all modalities. It also requires the ability to shape or regulate the activation of task- and time-relevant information in order to produce flexible and appropriate behavior. Some kind of regulatory process is critical: we store a wealth of information about the meanings of words and objects, but frequently only a subset of this knowledge is required for a task—indeed, other aspects of knowledge may actually be inappropriate and unhelpful. As an example, consider the radically different uses of the same knife in making; a sandwich: these can include piercing a package, slicing bread, meat, or cheese; scooping and spreading mustard or mayonnaise on the sandwich; and so on. Specific aspects of the knife's properties (and ways of holding and manipulating it) must be brought to the fore, one by one, while the most commonly listed property of cutting has to be inhibited in many of these activities. Indeed, in the case of scooping, the canonical function of the knife has to be disregarded altogether and replaced by a substituted function in place of another object (spoon). In sum, in addition to the acquisition and activation of conceptual knowledge, the ability to regulate and shape is critical to any complete account of semantic cognition.

The distinction between semantic representations and control processes that regulate processes acting on these representations helps to resolve a puzzle highlighted by a comparison of different, semantically impaired patient groups (i.e., patients who fail both verbal and nonverbal semantic tasks). Patients with ATL damage are not the only ones to exhibit poor semantic performance across different modalities. Indeed, it is possible to find a subset of aphasic patients who have multimodal semantic impairments (Chertkow, Bub, Deaudon, & Whitehead, 1997; Jefferies & Lambon Ralph, 2006) arising from temporoparietal or prefrontal damage rather than ATL damage. We refer to this pattern as *semantic aphasia* (SA) (Jefferies, Patterson, & Lambon Ralph, 2008). By directly comparing semantic aphasia and semantic dementia, we have been able to demonstrate that each group's failure on semantic tasks is qualitatively different and that they should not be considered as the same type of impairment.

We have hypothesized that the patient groups reflect two of the primary ingredients in semantic cognition: semantic dementia reflects a degradation of the core conceptual knowledge, whereas semantic aphasia arises from a deficit in the regulation of semantic cognition. We find that SD patients are highly consistent across different semantic tasks: patients who retained knowledge of an item in one task were typically able to demonstrate this knowledge in all other tasks. In contrast, SA patients show significant correlations/consistency only between different versions of the same semantic task (e.g., judgments of semantic association for words and pictures). Unlike SD, the SA patients' ability to retrieve information is inconsistent when tested across tasks with different semantic control demands (e.g., judgments of semantic association versus word-picture matching). Moreover, SA patients' ability to make semantic judgments can be predicted by how readily the relevant semantic dimension can be discerned and competitors rejected. For such patients, cues or constraints provided by the examiner can boost their performance considerably. The patients' errors in picture naming provide a further basis for differentiating the groups. The SD patients make frequent coordinate and superordinate semantic errors (such as saying "dog" or "animal" for goat). The SA patients also make associative errors (e.g., producing the response "nuts" for squirrel); these responses are virtually never seen in SD. These errors suggest that the SA patients retain a considerable amount of knowledge about unnamed targets (in order to be able to generate such errors) and suggest that their difficulty lies in directing activation toward the correct name and away from irrelevant associations.

These patient studies provide a direct convergence with fMRI studies of semantic processing in normal participants. These studies consistently implicate prefrontal cortex and the temporoparietal junction in tasks requiring controlled

semantic processing—for example, when a particular aspect of meaning must be selected or when there is strong competition from alternative responses (Thompson-Schill, Desposito, Aguirre, & Farah, 1997; Wagner, Pare-Blagoev, Clark, & Poldrack, 2001). It is possible that the relevant control or shaping processes may underpin executive/attentional functions more generally, as these same regions are commonly activated in a variety of tasks requiring cognitive control (Garavan et al., 2000; Peers et al., 2005). In keeping with this hypothesis, the SA patients tend to fail executive/attentional tasks even when they do not involve semantic information (Jefferies & Lambon Ralph, 2006).

Conclusion

This article has reviewed research spanning a wide range of research approaches. Behavioral investigations of developing children and neuropsychological patients, computational modeling investigations, and investigations of brain activity in healthy human volunteers using noninvasive imaging methods and TMS have all been used to support an overall account of the nature of semantic knowledge, its development and disintegration, and its instantiation in networks of interconnected areas of the brain. The approach has had some success in linking research from all these different methods under a common theoretical framework based on the principles of parallel distributed processing and in providing the stimulus for a considerable body of ongoing research. More work needs to be done to flesh out the theory and to better understand how activation of semantic and other forms of knowledge thought to depend on the anterior temporal lobes is influenced by activations in other brain areas.

REFERENCES

Benedet, M., Patterson, K., Gomez-Pastor, I., & de la Rocha, M. L. G. (2006). "Non-semantic" aspects of language in semantic dementia: As normal as they're said to be? *Neurocase*, *12*, 15–26.

Blazely, A., Coltheart, M., & Casey, B. J. (2005). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cogn. Neuropsychol.*, *22*, 695–717.

Bozeat, S., Lambon Ralph, M. A., Graham, K. S., Patterson, K., Wilkin, H., Rowland, J., et al. (2003). A duck with four legs: Investigating the structure of conceptual knowledge using picture drawing in semantic dementia. *Cogn. Neuropsychol.*, *20*, 27–47.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38*, 1207–1215.

Caramazza, A. (1997). How many levels of processing are there? *Cogn. Neuropsychol.*, *14*(1), 177–208.

Chertkow, H., Bub, D., Deaudon, C., & Whitehead, V. (1997). On the status of object concepts in aphasia. *Brain Lang.*, *58*, 203–232.

Cipolotti, L., & Warrington, E. K. (1995). Semantic memory and reading abilities: A case report. *J. Int. Neuropsychol. Soc.*, *1*, 104–110.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. J., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.*, *108*, 204–256.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., et al. (2000). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *NeuroImage*, *11*, 589–600.

Dilkina, K., McClelland, J. L., & Plaut, D. C. (2008). A single-system account of semantic and lexical deficits in five semantic dementia patients. *Cogn. Neuropsychol.*, *25*(2), 136–164.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *J. Pers. Soc. Psychol.*, *87*, 876–963.

Funnell, E. (1996). Response biases in oral reading: An account of the co-occurrence of surface dyslexia and semantic dementia. *Q. J. Exp. Psychol. A*, *49*(2), 417–446.

Garavan, H., Ross, T. J., Li, S. J., & Stein, E. A. (2000). A parametric manipulation of central executive functioning. *Cereb. Cortex*, *10*, 585–592.

Gelman, R., & Williams, E. M. (1998). Enabling constraints for cognitive development and learning: A domain specific epigenetic theory. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology*, Vol. 2: *Cognition, perception and development* (5th ed., pp. 575–630). New York: John Wiley and Sons.

Giovagnoli, A. R., Erbetta, A., Villani, F., & Avanzini, G. (2005). Semantic memory in partial epilepsy: Verbal and nonverbal deficits and neuroanatomical relationships. *Neuropsychologia*, *43*, 1482–1492.

Graham, N. L., Patterson, K., & Hodges, J. R. (2000). The impact of semantic memory impairment on spelling: Evidence from semantic dementia. *Neuropsychologia*, *38*, 143–163.

Hammers, A., Koepp, M. J., Richardson, M. P., Hurlemann, R., Brooks, D. J., & Duncan, J. S. (2003). Grey and white matter flumazenil binding in neocortical epilepsy with normal MRI. A PET study of 44 patients. *Brain*, *126*, 1300–1318.

Hermann, B., Davies, K., Foley, K., & Bell, B. (1999). Visual confrontation naming outcome after standard left anterior temporal lobectomy with sparing versus resection of the superior temporal gyrus: A randomized prospective clinical trial. *Epilepsia*, *40*, 1070–1076.

Hickok, G., & Poeppel, D. (2007). Opinion—The cortical organization of speech processing. *Nat. Rev. Neurosci.*, *8*, 393–402.

Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 161–187). Hillsdale, NJ: Erlbaum.

Hinton, G. E. (1989). Learning distributed representations of concepts. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 46–61). Oxford, UK: Clarendon Press.

Hodges, J. R., Graham., N., & Patterson, K. (1995). Charting the progression of semantic dementia: Implications for the organisation of semantic memory. *Memory*, *3*, 463–495.

Hodges, J. R., & Patterson, K. (2007). Semantic dementia: A unique clinicopathological syndrome. *Lancet Neurol.*, *6*, 1004–1014.

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, *115*, 1783–1806.

Hodges, J. R., Salmon, D. P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia*, *30*, 301–314.

Howard, D., & Patterson, K. (1992). *The Pyramids and Palm Trees Test: A test of semantic access from words and pictures*. Bury St. Edmunds, UK: Thames Valley Test Company.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia vs. semantic dementia: A case-series comparison. *Brain*, *129*, 2132–2147.

Jefferies, E., Patterson, K., & Lambon Ralph, M. A. (2008). Deficits of knowledge vs. executive control in semantic cognition: Insights from cued naming. *Neuropsychologia*, *46*, 649–658.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: Harvard University Press.

Keil, F. (1991). The emergence of theoretical beliefs as constraints on concepts. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition*. Hillsdale, NJ: Erlbaum.

Lambon Ralph, M. A., Lowe, C., & Rogers, T. T. (2007). Neural basis of category-specific semantic deficits for living things: Evidence from semantic dementia, HSVE and a neural network model. *Brain*, *130*, 1127–1137.

Lambon Ralph, M. A., McClelland, J. L., Patterson, K., Galton, C. J., & Hodges, J. R. (2001). No right to speak? The relationship between object naming and semantic impairment: Neuropsychological evidence and a computational model. *J. Cogn. Neurosci.*, *13*, 341–356.

Lambon Ralph, M. A., & Patterson, K. (2008). Generalisation and differentiation in semantic memory: Insights from semantic dementia. *Ann. NY Acad. Sci.*, *1124*, 61–76.

Lambon Ralph, M. A., Pobric, G., & Jefferies, E. (2009). Conceptual knowledge is underpinned by the temporal pole bilaterally: Novel data from rTMS. *Cereb. Cortex*, *19*, 832–838.

Levelt, W. J. M. (1989). *Speaking, from intention to articulation*. Cambridge, MA: MIT Press.

Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cogn. Dev.*, *8*, 291–318.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, *38*, 487–497.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25–45.

Martin, A., & Chao, L. L. (2001). Semantic memory in the brain: Structure and processes. *Curr. Opin. Neurobiol.*, *11*, 194–201.

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 8–45). New York: Oxford University Press.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.*, *102*, 419–457.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nat. Rev. Neurosci.*, *4*, 310–322.

Mervis, C. B. (1987). Child basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge, UK: Cambridge University Press.

Noppeney, U., Patterson, K., Tyler, L. K., Moss, H., Stamatakis, E. A., Bright, P., et al. (2007). Temporal lobe lesions and semantic impairment: A comparison of herpes simplex virus encephalitis and semantic dementia. *Brain*, *130*, 1138–1147.

O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Comput.*, *8*, 895–938.

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, *30*(12), 1025–1040.

Patterson, K., Lambon Ralph, M. A., Hodges, J. R., & McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia*, *39*, 709–724.

Patterson, K., Lambon Ralph, M. A., Jefferies, E., Woollams, A., Jones, R., Hodges, J. R., & Rogers, T. T. (2006). "Presemantic" cognition in semantic dementia: Six deficits in search of an explanation. *J. Cogn. Neurosci.*, *18*(2), 169–183.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.*, *8*, 976–987.

Pauen, S. (2002). The global-to-basic shift in infants' categorical thinking: First evidence from a longitudinal study. *Int. J. Behav. Dev.*, *26*(6), 492–499.

Peers, P. V., Ludwig, C. J. H., Rorden, C., Cusack, R., Bonfiglioli, C., Bundesen, C., et al. (2005). Attentional functions of parietal and frontal cortex. *Cereb. Cortex*, *15*, 1469–1484.

Pinker, S. (1991). Rules of language. *Science*, *253*, 530–535.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychol. Rev.*, *103*, 56–115.

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (submitted). Non-verbal semantic memory in the anterior temporal lobes: TMS evidence.

Pobric, G. G., Jefferies, E., & Lambon Ralph, M. A. (2007). Anterior temporal lobes mediate semantic representation: Mimicking semantic dementia by using rTMS in normal participants. *Proc. Natl. Acad. Sci. USA*, *104*, 20137–20141.

Powell, H. W. R., Parker, G. J. M., Alexander, D. C., Symms, M. R., Boulby, P. A., Wheeler-Kingshott, C. A. M., et al. (2007). Abnormalities of language networks in temporal lobe epilepsy. *NeuroImage*, *36*, 209–221.

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends Cogn. Sci.*, *6*, 416–421.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 227–270). Cambridge, MA: MIT Press.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). The structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychol. Rev.*, *111*, 205–235.

Rogers, T. T., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2004). Natural selection: The impact of semantic impairment on lexical and object decision. *Cogn. Neuropsychol.*, *21*(2/3/4), 331–352.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rogers T. T. & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *J. Exp. Psychol. Gen.*, *136*, 451–469.

Rumelhart, D. E. (1990). Brain style computation: Learning and generalization. In S. F. Zornetzer, J. L. Davis, & C. Lau (Eds.), *An introduction to neural and electronic networks* (pp. 405–420). San Diego: Academic Press.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 216–271). Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., and the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vols. 1 & 2). Cambridge, MA: MIT Press.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 3–30). Cambridge, MA: MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychol. Rev.*, *96*, 523–568.

Snowden, J. S., Goulding, P. J., & Neary, D. (1989). Semantic dementia: A form of circumscribed temporal atrophy. *Behav. Neurol.*, *2*, 167–182.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.*, *99*, 195–231.

Thompson-Schill, S. L., Desposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc. Natl. Acad. Sci. USA*, *94*, 14792–14797.

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. J. (1996). Functional-anatomy of a common semantic system for words and pictures. *Nature*, *383*, 254–256.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*, 329–338.

Ward, J., Stott, R., & Parkin, A. J. (2000). The role of semantics in reading and spelling: Evidence for the "summation hypothesis." *Neuropsychologia*, *38*, 1643–1653.

Warrington, E. K. (1975). Selective impairment of semantic memory. *Q. J. Exp. Psychol.*, *27*, 635–657.

Welbourne, S. R., & Lambon Ralph, M. A. (2005). Subtracting subtractivity? A connectionist account of recovery in single word reading following brain damage. *Cogn. Affect. Behav. Neurosci.*, *5*, 77–92.

Wilkins, A., & Moscovitch, M. (1978). Selective impairment of semantic memory after temporal lobectomy. *Neuropsychologia*, *16*, 73–79.

Wise, R. (2003). Language systems in normal and aphasic human subjects: Functional imaging studies and inferences from animal studies. *Br. Med. Bull.*, *65*, 95–119.

Woollams, A. M., Lambon Ralph, M. A., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychol. Rev.*, *114*(2), 316–339.

Woollams, A. M., Cooper-Pye, E., Hodges, J. R., & Patterson, K. (2008). Anomia: A doubly typical signature of semantic dementia. *Neuropsychologia. 46*, 2503–2514.

73 Two Views of Brain Function

MARCUS E. RAICHLE

ABSTRACT In this chapter I encourage a broader discussion of how we view brain function. Traditionally, experiments of those studying brain functions have largely focused on the manner in which it responds to the momentary demands of the environment. By their very nature such experiments encourage a primarily reflexive view of brain function. Although such an approach has been remarkably productive, it ignores the alternative possibility that the brain's operations are mainly intrinsic, involving the maintenance of information for interpreting, responding to, and even predicting environmental demands. Here I will argue that it is the latter view that best characterizes the nature of brain function.

Since the 19th century and possibly longer, two views of brain function have existed (for a brief historical review see chapter 1 in Llinas, 2001). One view posits that the brain is primarily reflexive, driven by the momentary demands of the environment. The other view is that the brain's operations are mainly intrinsic, involving the maintenance of information for interpreting, responding to, and even predicting environmental demands. The former has motivated most neuroscience research, including that with functional neuroimaging. This is likely the case because experiments designed to measure brain responses to controlled stimuli and carefully designed tasks can be rigorously controlled, whereas evaluating the behavioral relevance of intrinsic brain activity can be an elusive enterprise. I believe the successes in studying evoked activity have caused us to ignore the possibility that our experiments reveal only a small fraction of the actual functional activity performed by our brain. This has certainly been the case in cognitive neuroscience until quite recently.

It was the concept of a default mode of brain function (Gusnard & Raichle, 2001; Raichle & Snyder, 2007; Raichle et al., 2001), which resulted from our desire to understand activity decreases from a resting state during task performance, that caused us to take seriously the notion that intrinsic activity likely plays an important role in brain function. Two other observations reinforced this belief. First was the demonstration that "noise" in the fMRI BOLD signal, first noted by Bharat Biswal and colleagues (Biswal, Yetkin, Haughton, & Hyde, 1995) and now actively being studied by many investigators (for a recent review see M. Fox & Raichle, 2007), actually represents ongoing (intrinsic), spatially coherent activity within brain systems including the default mode network (Greicius, Krasnow, Reiss, & Menon, 2003). The second observation was that intrinsic functional activity accounts for the overwhelming majority of the brain's enormous energy budget (Raichle, 2006; Raichle & Mintun, 2006). In this chapter I review how these observations converged to underscore the importance of the brain's intrinsic activity for me.

MARCUS E. RAICHLE Washington University School of Medicine, St. Louis, Missouri

The origins of an idea

By the early 1980s, positron emission tomography (PET) began to receive serious attention as a potential functional neuroimaging device in human subjects. (For a detailed historical account see Raichle, 2000.) The study of human cognition with neuroimaging was aided greatly in the 1980s by the involvement of cognitive psychologists whose experimental strategies for dissecting human behaviors fit well with the emerging capabilities of functional brain imaging (Posner & Raichle, 1994). Subtracting functional images acquired in a task state from ones acquired in a control state was a natural extension of mental chronometry (Posner, 1986), in which one measures the time required to complete specific mental operations isolated by the careful selection of task and control states. This approach, in various forms, has dominated the cognitive neuroscience agenda ever since with remarkably productive results.

For the better part of a decade following the introduction of subtractive methodology to neuroimaging, the vast majority of changes reported in the literature were activity increases, or *activations*, as they were almost universally called. Activity increases but not decreases are expected in subtractions of a control condition from a task condition as long as the *assumption of pure insertion* is valid. To illustrate, using an example based on mental chronometry, say that one's control task requires a key press to a simple stimulus such as the appearance of a point of light in the visual field, whereas the task state of interest requires a decision about the color of the light prior to the key press. Assuming pure insertion, the response latency difference between conditions is interpretable as the time needed to perform color discrimination. However, the time needed to press a key might be affected by the nature of the decision process itself, violating the assumption of pure insertion. More generally, the brain state

underlying any action could have been altered by the introduction of an additional process. Interestingly, functional neuroimaging helped address the question of pure insertion by employing the device of *reverse subtraction*. Thus in certain circumstances subtracting task-state data from control-state data revealed negative responses, or *task-specific deactivations* (for examples and further discussion of this interesting issue see S. Petersen, van Mier, Fiez, & Raichle, 1998; Raichle, 1998; Raichle et al., 1994). It was clearly shown, just as psychologists had suspected, that processes active in a control state could be modified when paired with a particular task. However, none of this work prepared us for nor anticipated "the problem."

"The problem," as we now think of it, arose unexpectedly when we noted, quite by accident, that activity decreases were present in our subtraction images even when the control state was either visual fixation or eyes-closed rest. What particularly caught our attention was the fact that, regardless of the task under investigation, the activity decreases almost always included the posterior cingulate and adjacent precuneus, a region we nicknamed MMPA for "medial mystery parietal area."

The first formal characterization of task-induced activity decreases was a large meta-analysis of published PET data from our group (G. Shulman et al., 1997). This study generated a set of iconic images, now generally referred to as the *default mode network* (figure 73.1*A*) after our later paper on the subject (Raichle et al., 2001). The unique identity of this network was amply confirmed in later meta-analyses by Jeffery Binder and colleagues at the Medical College of Wisconsin (Binder et al., 1999) and Bernard Mazoyer and his colleagues (Mazoyer et al., 2001) in France. Similar observations are now an everyday occurrence in laboratories throughout the world, leaving little doubt that a specific set of brain areas decrease their activity across a remarkably wide array of task conditions when compared to a passive control condition such as visual fixation.

The finding of a network of brain areas frequently seen to decrease its activity during goal-directed tasks (figure 73.1*A*) was both surprising and challenging. Surprising because the areas involved had not previously been recognized as a system in the same way we might think of the motor or visual system. And challenging because initially it was unclear how to characterize their activity in a passive or resting condition.

For us the issue of characterizing activity decreases came to a head in 1998 when a paper we were attempting to publish was rejected because of the way in which we characterized activity changes of the type seen in figure 73.1*A*. One of the referees wrote, "This is the most controversial aspect of this paper as it (1) cannot be ruled out that these signal changes are actual activations in the so-called resting state and (2) the physiological mechanisms underpinning a genuine BOLD signal decrease remain a matter of speculation."

It was clear that we needed a way to determine whether or not task-induced activity decreases were simply "activations" present in the absence of an externally directed task and why they should appear in both PET and fMRI functional neuroimaging studies. In pursuing this question we employed quantitative PET measurements of regional brain blood flow and oxygen consumption to define a physiologic baseline. The details of this work have been recounted on several occasions (Gusnard & Raichle, 2001; Raichle & Snyder, 2007; Raichle et al., 2001; Raichle & Mintun, 2006), including the third edition of *The Cognitive Neurosciences* (Gusnard & Raichle, 2004), and thus will not be repeated here. Suffice to say that this work allowed moving forward on the assumption that activity within the default mode network did not represent conventional activations in the resting state.

Having arrived at the view that the brain has a default mode of function through our analysis of activity decreases, we began to take seriously claims that there was likely much more to brain function than that revealed by momentary demands of the environment. Two bodies of information have been especially persuasive.

The cost of intrinsic activity

One of the challenges we face is trying to adjudicate the relative importance of intrinsic and evoked activity. For me the most persuasive approach has been to consider their relative costs. Let us turn then to a brief discussion of the budgeting of brain energy consumption.

In the average adult human, the brain represents about 2% of the total body weight but about 20% of the energy consumed (Clark & Sokoloff, 1999), 10 times that predicted by its weight alone. Relative to this very high rate of ongoing or "basal" metabolism (usually measured while resting quietly awake with eyes closed), the amount dedicated to regional imaging signals is remarkably small.

The regional increases in absolute blood flow associated with imaging signals as measured with PET are usually no more than 5–10% of the resting blood flow of the brain. These are modest modulations in ongoing circulatory activity that rarely affect the overall rate of brain blood flow during even the most arousing perceptual and vigorous motor activity (P. Fox, Burton, & Raichle, 1987; Friston et al., 1990; Lennox, 1931; Madsen et al., 1995; Roland, Eriksson, Stone-Elander, & Widen, 1987; Sokoloff, Mangold, Wechsler, Kennedy, & Kety, 1955).

The modest nature of these task-induced increases in blood flow is further underscored when one considers the increase in energy consumption they represent. Recall that the average resting metabolic activity of the brain is sup-

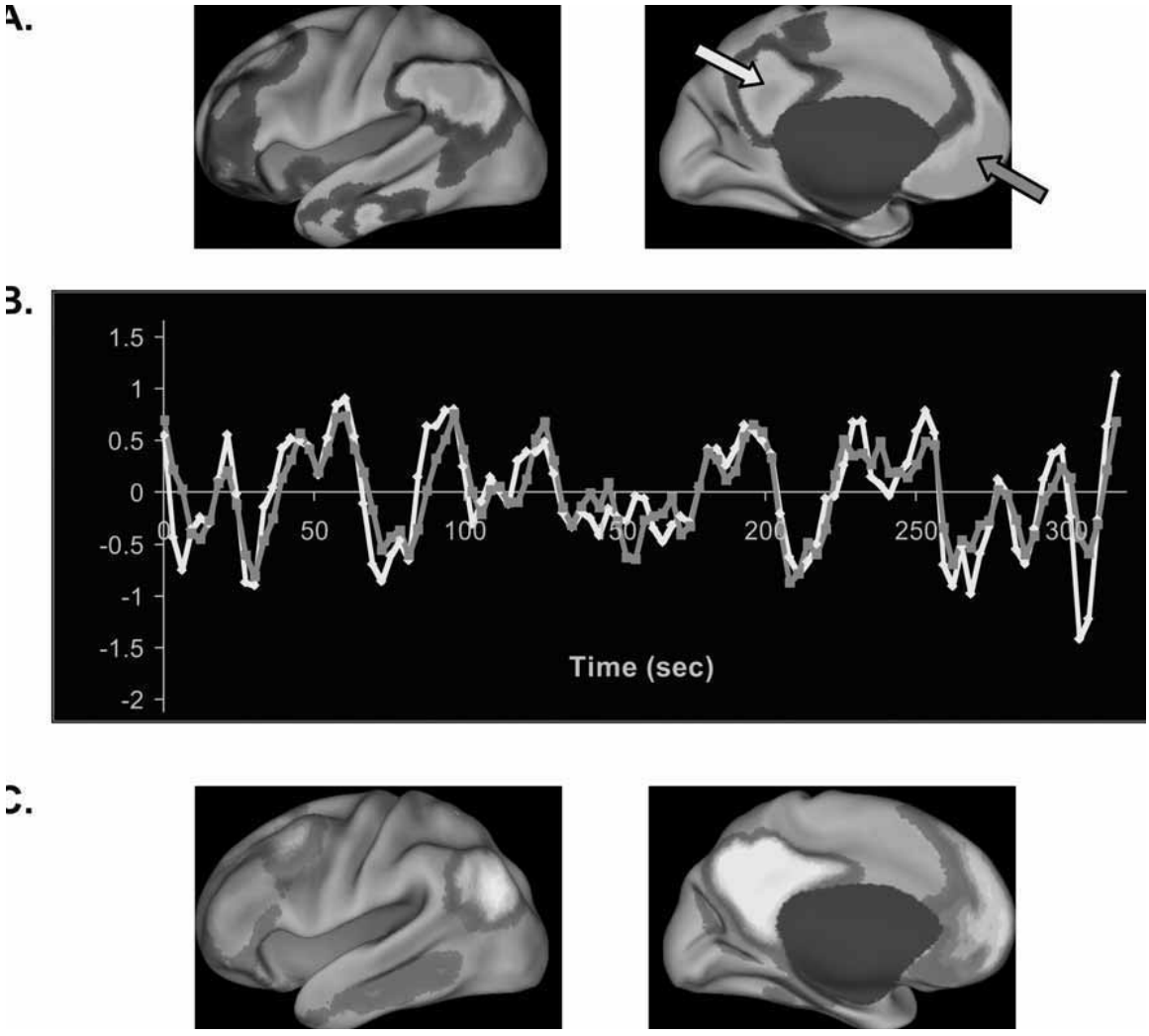


FIGURE 73.1 Performance of a wide variety of tasks has called attention to a group of brain areas (*A*) that decrease their activity during task performance (data adapted from G. Shulman et al., 1997). These areas are often referred to as the brain's default mode network after our initial work on them (Raichle et al., 2001). If one records the spontaneous fMRI BOLD signal activity in these areas in the resting state (arrows, *A*) what emerges is a remarkable similarity in the behavior of the signals between areas (*B*), a phenomenon originally described by Biswal and colleagues (1995) in the somatomotor cortex and later in the default mode network by Greicius and colleagues (2003). Using these fluctuations to analyze the network as a whole (M. Fox et al., 2005; Vincent et al., 2006) reveals a level of functional organization (*C*) that parallels that seen in the task-related activity decreases. These data provide a dramatic demonstration that the ongoing organization of the human brain likely provides a critical context for all human behaviors. (These data were adapted from our earlier published work: M. Fox et al., 2005; Gusnard & Raichle, 2001; Raichle et al., 2001; G. Shulman et al., 1997.) (See color plate 87.)

ported by the nearly complete (>90%) oxidation of glucose to carbon dioxide and water, producing approximately 32 moles of ATP per mole of glucose consumed (Siesjö, 1978). Imaging signal activations, however, are associated with increases in glucose utilization that are not accompanied by a proportionate increase in oxygen consumption (Blomqvist et al., 1994; P. Fox, Raichle, Mintun, & Dence, 1988; Madsen et al., 1995), resulting in the production of only 2 moles of ATP per mole of glucose consumed, typical of glycolysis. Estimates of the actual increases in oxygen consumption vary somewhat (P. Fox & Raichle, 1986; P. Fox et al., 1988; Fujita, Kuwahara, Reutens, & Gjedde, 1999; Mintun, Vlassenko, Shulman, & Snyder, 2002; Roland, Eriksson, Widen, & Stone-Elander, 1989) but are always less than that predicted by the increase in blood flow. From knowledge of these relationships, one can estimate that if blood flow and glucose utilization increase by 10%, but oxygen consumption does not, the local energy consumption increase due to a typical task-related response could be as little as 1%. It becomes clear, then, that the brain continuously expends a considerable amount of energy even in the absence of a particular task (i.e., when a subject is awake and at "rest").

What is the nature of this intrinsic activity present even at rest that commands such a large amount of the brain's energy resources?

It is tempting to assume that this intrinsic or resting-state energy utilization reflects simple housekeeping functions such as neuronal repair or protein trafficking. However, the preponderance of evidence suggests that such functions consume a relatively small fraction of the brain's energy budget. Measurements of brain energy metabolism using magnetic resonance spectroscopy (R. Shulman, Hyder, & Rothman, 2001; R. Shulman, Rothman, Behar, & Hyder, 2004; Sibson et al., 1997, 1998) in a variety of experimental settings have indicated that up to 80% of the entire energy consumption of the brain at rest is devoted to glutamate cycling and, hence, neural signaling processes. Complementary analyses using extant anatomic, physiologic, and metabolic data (Ames, 2000; Attwell & Laughlin, 2001; Lennie, 2003; Wong-Riley, 1989) to assess the cost of different components of excitatory signaling in the gray matter have arrived at similar conclusions. Such estimates leave for future consideration the demands placed on the brain's energy budget by the functional activity of inhibitory interneurons (Ackerman, Finch, Babb, & Engel, 1984; Buzsaki, Kaila, & Raichle, 2007; Chatton, Pellerin, & Magistretti, 2003; McCasland & Hibbard, 1997; Patel et al., 2005; Waldvogel et al., 2000). That evidence notwithstanding, it is likely to remain the case that a significant fraction of the energy consumed by the brain (quite possibly the majority) is due to functionally significant intrinsic neuronal activity.

From this cost-based analysis of brain functional activity it seems reasonable to conclude that intrinsic activity may be as significant as, if not more than, evoked activity in terms of overall brain function. Taking this position converts one's view of the brain as a system primarily responding to changing contingencies to one operating on its own, intrinsically, with sensory information interacting with rather than determining the operation of the system. This view has historical (Llinas, 1988) and recent theoretical (Olshausen & Field, 2005), as well as experimental, support (Arieli, Sterkin, Grinvald, & Aertsent, 1996; Fiser, Chiu, & Weliky, 2004; Kenet, Bibitchkov, Tsodyks, Grinvald, & Arieli, 2003). It seems highly likely that ultimate constraints on the behaviors we study with neuroimaging or any other technique will be significantly determined by this intrinsic activity.

The organization of intrinsic activity

A second body of evidence that persuades us to take seriously the importance of intrinsic activity is its remarkable degree of organization. For us this organization was first revealed in the activity decreases we and others observed in our studies with functional neuroimaging (figure 73.1*A*). More striking, however, have been the patterns of activity revealed in the analysis of the "noise" in the fMRI BOLD signal when subjects are resting quietly in the scanner with their eyes closed or simply maintaining visual fixation.

A prominent feature of fMRI is that the unaveraged signal is quite noisy (figure 73.1*B*), prompting researchers to average their data to reduce this "noise" and increase the signals they seek. As first shown by Biswal, Hudetz, Yetkin, Haughton, and Hyde (1997), a considerable fraction of the variance in the BOLD signal in the frequency range below 0.1 Hz appears to reflect spontaneous fluctuating neuronal activity that exhibits striking patterns of coherence within known brain systems (figure 73.1*C*), even in the absence of observable behaviors associated with those systems (for a recent review of this rapidly expanding literature see M. Fox & Raichle, 2007). Additionally, these patterns of coherence are remarkably consistent among individuals as well as across subject groups. The value of studying these resting-state BOLD fluctuations has been well articulated (Buckner & Vincent, 2007). But what does intrinsic activity represent?

One possibility is that intrinsic activity simply represents unconstrained, spontaneous cognition—our daydreams or, more technically, stimulus-independent thoughts (SITs; Antrobus, 1968; Mason et al., 2007; McGuire, Paulesu, Frackowiak, & Frith, 1996). But from a cost perspective SITs are highly unlikely to account for more energy consumption than that elicited by responding to controlled stimuli, which accounts for a very small fraction of total brain activity (Raichle & Mintun, 2006). Most telling is the recent observations that spatially coherent, spontaneous BOLD activity is present under general anesthesia (Vincent et al., 2007) and during sleep (Fukunaga et al., 2006). These two important observations suggest that intrinsic activity cannot simply be a reflection of conscious mental activity. Rather, it likely reflects a more fundamental property of brain functional organization.

Among the possible functions of this intrinsic activity is the regulation of neuronal responsiveness. Neurons continuously receive both excitatory and inhibitory inputs. The "balance" of these stimuli determines the responsiveness (or gain) of neurons to correlated inputs and, in so doing, potentially sculpts communication pathways in the brain (Abbott & Chance, 2005; Haider, Duque, Hasenstaub, & McCormick, 2006; Laughlin & Sejnowski, 2003; Salinas & Sejnowski, 2001). Balance also manifests at a large-systems level. For example, neurologists know that strokes damaging cortical centers controlling eye movements lead to deviation of the eyes toward the side of the lesion, implying the preexisting presence of "balance." Another well-known example first demonstrated in the visual system of the cat is the "Sprague effect" (Sprague, 1966). It may be that in the normal brain, a balance of opposing forces enhances

the precision of a wide range of processes. Thus "balance" might be viewed as a necessary enabling, but costly, element of brain function.

A more expanded view is that intrinsic activity instantiates, in ways yet to be fully understood, the maintenance of information for interpreting, responding to, and even *predicting* environmental demands. In this regard, a useful conceptual framework from theoretical neuroscience posits that the brain operates as a Bayesian inference engine designed to generate predictions about the future (Kersten, Mamassian, & Yuille, 2004; Knill & Pouget, 2004; Olshausen, 2003). Beginning with a set of "advance" predictions at birth, the brain is then sculpted by worldly experience to represent intrinsically a "best guess" ("priors" in Bayesian parlance) about the environment and, in the case of humans at least, to make predictions about the future. This is a theme presciently enunciated many years ago by the late David Ingvar in his memorable essay "Memory of the Future" (Ingvar, 1985).

An important question for future research is how to incorporate studies of intrinsic brain activity into an already busy program of work devoted to evoked activity. Some, of course, will elect not to do so. However, limiting one's approach in this way will eventually limit its potential if it is not nourished by a broader consideration and understanding of such relevant neurobiology. What is required is an expanded framework upon which to base one's research agenda. Neuroscience and the behavioral sciences *together* must provide that framework, which is one that we heartily endorse.

Cognitive neuroscientists for their part will need to become more familiar with a broad range of approaches to the study of spontaneous activity of neurons (Arieli et al., 1996; Buzsaki & Draguhn, 2004; Foster & Wilson, 2006; Kay, 2005; Kenet et al., 2003; Leopold, Murayama, & Logothetis, 2003), which can include work in humans (He, Snyder, Zempel, Smyth, & Raichle, 2008). In this regard, descriptions of slow fluctuations (nominally, <0.1 Hz) in neuronal membrane polarization—so-called up and down states—are intriguing (Hahn, Sakmann, & Mehta, 2006; Isomura et al., 2006; Luczak, Bartho, Marguet, Buzsaki, & Harris, 2006; C. Petersen, Hahn, Mehta, Grinvald, & Sakmann, 2003). Not only does their temporal frequency approach that of the spontaneous fluctuations in the fMRI BOLD signal along with other low-frequency fluctuations (He et al.), but their functional consequences may be relevant to an understanding of the variability in task-evoked brain activity (Arieli et al.; M. Fox, Snyder, Zacks, & Raichle, 2006), as well as behavioral variability in human performance (M. Fox, Snyder, Vincent, & Raichle, 2007; Gilden, Thornton, & Mallon, 1995). Neuroscientists for their part need to be aware of the expanded view of intrinsic activity afforded by neuroimaging and the potential to relate this not only to their own work at the cellular level but also to the behavior we all seek to understand that is instantiated in the large-scale organization of the brain's intrinsic activity.

REFERENCES

Abbott, L. F., & Chance, F. S. (2005). Drivers and modulators from push-pull and balanced synaptic input. *Prog. Brain Res.*, *149*, 147–155.

Ackerman, R. F., Finch, D. M., Babb, T. L., & Engel, J. J. (1984). Increased glucose utilization during long-duration recurrent inhibition of hippocampal pyramidal cells. *J. Neurosci.*, *4*, 251–264.

Ames, A. I. (2000). CNS energy metabolism as related to function. *Brain Res. Brain Res. Rev.*, *34*, 42–68.

Antrobus, J. S. (1968). Information theory and stimulus-independent thought. *Br. J. Psychol.*, *59*(4), 423–430.

Arieli, A., Sterkin, A., Grinvald, A., & Aertsen, A. (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science*, *273*(5283), 1868–1871.

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.*, *21*, 1133–1145.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state: A functional MRI study. *J. Cogn. Neurosci.*, *11*(1), 80–95.

Biswal, B., Hudetz, A. G., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1997). Hypercapnia reversibly suppresses low-frequency fluctuations in the human motor cortex during rest using echo-planar MRI. *J. Cereb. Blood Flow Metab.*, *17*(3), 301–308.

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.*, *34*(4), 537–541.

Blomqvist, G., Seitz, R. J., Sjogren, I., Halldin, C., Stone-Elander, S., Widen, L., et al. (1994). Regional cerebral oxidative and total glucose consumption during rest and activation studied with positron emission tomography. *Acta Physiol. Scand.*, *151*, 29–43.

Buckner, R. L., & Vincent, J. L. (2007). Unrest at rest: Default activity and spontaneous network correlations. *NeuroImage*, *37*(4), 1091–1096; discussion, 1097–1099.

Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926–1929.

Buzsaki, G., Kaila, K., & Raichle, M. (2007). Inhibition and brain work. *Neuron*, *56*(5), 771–783.

Chatton, J.-Y., Pellerin, L., & Magistretti, P. J. (2003). GABA uptake into astrocytes is not associated with significant metabolic cost: Implications for brain imaging of inhibitory transmission. *Proc. Natl. Acad. Sci. USA*, *100*, 12456–12461.

Clark, D. D., & Sokoloff, L. (1999). Circulation and energy metabolism of the brain. In G. J. Siegel, B. W. Agranoff, R. W. Albers, S. K. Fisher, & M. D. Uhler (Eds.), *Basic neurochemistry: Molecular, cellular and medical aspects* (6th ed., pp. 637–670). Philadelphia: Lippincott-Raven.

Fiser, J., Chiu, C., & Weliky, M. (2004). Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature*, *431*, 573–578.

Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, *440*(7084), 680–683.

Fox, M. D., & Raichle, M. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.*, *8*, 700–711.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. USA*, *102*(27), 9673–9678.

Fox, M. D., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2007). Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron*, *56*(1), 171–184.

Fox, M. D., Snyder, A. Z., Zacks, J. M., & Raichle, M. E. (2006). Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. *Nat. Neurosci.*, *9*(1), 23–25.

Fox, P. T., Burton, H., & Raichle, M. E. (1987). Mapping human somatosensory cortex with positron emission tomography. *J. Neurosurg.*, *67*, 34–43.

Fox, P. T., & Raichle, M. E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc. Natl. Acad. Sci. USA*, *83*, 1140–1144.

Fox, P. T., Raichle, M. E., Mintun, M. A., & Dence, C. (1988). Nonoxidative glucose consumption during focal physiologic neural activity. *Science*, *241*, 462–464.

Friston, K. J., Frith, C. D., Liddle, P. F., Dolan, R. J., Lammertsma, A. A., & Frackowiak, R. S. J. (1990). The relationship between global and local changes in PET scans. *J. Cereb. Blood Flow Metab.*, *10*, 458–466.

Fujita, H., Kuwahara, H., Reutens, O., & Gjedde, A. (1999). Oxygen consumption of cerebral cortex fails to increse during continued vibrotactile stimulation. *J. Cereb. Blood Flow Metab.*, *19*, 266–271.

Fukunaga, M., Horovitz, S. G., van Gelderen, P., de Zwart, J. A., Jansma, J. M., Ikonomidou, V. N., et al. (2006). Large-amplitude, spatially correlated fluctuations in BOLD fMRI signals during extended rest and early sleep stages. *Magn. Reson. Imaging*, *24*(8), 979–992.

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/f noise in human cognition. *Science*, *267*(5205), 1837–1839.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. USA*, *100*(1), 253–258.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nat. Rev. Neurosci.*, *2*(10), 685–694.

Gusnard, D. A., & Raichle, M. (2004). Functional imaging, neurophysiology and the resting state of the human brain. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences* (3rd ed., pp. 1267–1280). Cambridge, MA: MIT Press.

Hahn, T. T., Sakmann, B., & Mehta, M. R. (2006). Phase-locking of hippocampal interneurons' membrane potential to neocortical up-down states. *Nat. Neurosci.*, *9*(11), 1359–1361.

Haider, B., Duque, A., Hasenstaub, A. R., & McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.*, *26*(17), 4535–4545.

He, B. J., Snyder, A. Z., Zempel, J. M., Smyth, M., & Raichle, M. (2008). Electrophysiological correlates of the brain's intrinsic large-scale functional architecture. *Proc. Natl. Acad. Sci. USA*, *105*, 16039–16044.

Ingvar, D. H. (1985). "Memory of the future": an essay on the temporal organization of conscious awareness. *Hum. Neurobiol.*, *4*(3), 127–136.

Isomura, Y., Sirota, A., Ozen, S., Montgomery, S., Mizuseki, K., Henze, D. A., et al. (2006). Integration and segregation of activity in entorhinal-hippocampal subregions by neocortical slow oscillations. *Neuron*, *52*(5), 871–882.

Kay, L. M. (2005). Theta oscillations and sensorimotor performance. *Proc. Natl. Acad. Sci. USA*, *102*, 3863–3868.

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., & Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature*, *425*(6961), 954–956.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol.*, *55*, 271–304.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci.*, *27*(12), 712–719.

Laughlin, S. B., & Sejnowski, T. J. (2003). Communication in neuronal networks. *Science*, *301*(5641), 1870–1874.

Lennie, P. (2003). The cost of cortical computation. *Curr. Biol.*, *13*, 493–497.

Lennox, W. G. (1931). The cerebral circulation. XV. Effect of mental work. *Arch. Neurol. Psychiatry*, *26*, 725–730.

Leopold, D. A., Murayama, Y., & Logothetis, N. K. (2003). Very slow activity fluctuations in monkey visual cortex: Implications for functional brain imaging. *Cereb. Cortex*, *13*(4), 422–433.

Llinas, R. (1988). The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. *Science*, *242*, 1654–1664.

Llinas, R. (2001). *I of the vortex*. Cambridge, MA: MIT Press.

Luczak, A., Bartho, P., Marguet, S. L., Buzsaki, G., & Harris, K. D. (2006). Sequential structure of neocortical spontaneous activity in vivo. *Proc. Natl. Acad. Sci. USA*, *104*, 347–352.

Madsen, P. L., Hasselbalch, S. G., Hagemann, L. P., Olsen, K. S., Bulow, J., Holm, S., et al. (1995). Persistent resetting of the cerebral oxygen/glucose uptake ratio by brain activation: Evidence obtained with the Kety-Schmidt technique. *J. Cereb. Blood Flow Metab.*, *15*, 485–491.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science*, *315*(5810), 393–395.

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houde, O., et al. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res. Bull.*, *54*(3), 287–298.

McCasland, J. S., & Hibbard, L. S. (1997). GABAergic neurons in barrel cortex show strong, whisker-dependent metabolic activation during normal behavior. *J. Neurosci.*, *17*, 5509–5527.

McGuire, P. K., Paulesu, E., Frackowiak, R. S., & Frith, C. D. (1996). Brain activity during stimulus independent thought. *NeuroReport*, *7*(13), 2095–2099.

Mintun, M., Vlassenko, A. G., Shulman, G. I., & Snyder, A. Z. (2002). Time-related increase of oxygen utilization in continuously activated human visual cortex. *NeuroImage*, *16*, 531–537.

Olshausen, B. A. (2003). Principles of image representation in visual cortex. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1603–1615). Cambridge, MA: MIT Press.

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding V1? *Neural Comput.*, *17*, 1665–1699.

Patel, A. B., de Graaf, R. A., Mason, G. F., Rothman, D. L., Shulman, R. G., & Behar, K. L. (2005). The contribution of GAGA to glutamate/glutamine cycling and energy metabolism in the rat cortex in vivo. *Proc. Natl. Acad. Sci. USA*, *102*, 5588–5593.

Petersen, C. C., Hahn, T. T., Mehta, M., Grinvald, A., & Sakmann, B. (2003). Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. *Proc. Natl. Acad. Sci. USA*, *100*(23), 13638–13643.

Petersen, S. E., van Mier, H., Fiez, J. A., & Raichle, M. E. (1998). The effects of practice on the functional anatomy of task performance. *Proc. Natl. Acad. Sci. USA*, *95*(3), 853–860.

Posner, M. (1986). *Chronometric explorations of mind*. New York: Oxford University Press.

Posner, M. J., & Raichle, M. E. (1994). *Images of mind*. New York: Scientific American Library.

Raichle, M. E. (1998). Behind the scenes of functional brain imaging: A historical and physiological perspective. *Proc. Natl. Acad. Sci. USA*, *95*(3), 765–772.

Raichle, M. (2000). A brief history of human functional brain mapping. In A. Toga & J. Mazziotta (Eds.), *Brain mapping: The systems* (pp. 33–75). San Diego: Academic Press.

Raichle, M. E. (2006). The brain's dark energy. *Science*, *314*(5803), 1249–1250.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. M., Pardo, J. V., Fox, P. T., et al. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb. Cortex*, *4*(1), 8–26.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. USA*, *98*(2), 676–682.

Raichle, M. E., & Mintun, M. A. (2006). Brain work and brain imaging. *Annu. Rev. Neurosci.*, *29*, 449–476.

Raichle, M., & Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage*, *37*, 1083–1090.

Roland, P. E., Eriksson, L., Stone-Elander, S., & Widen, L. (1987). Does mental activity change the oxidative metabolism of the brain? *J. Neurosci.*, *7*, 2373–2389.

Roland, P. E., Eriksson, L., Widen, L., & Stone-Elander, S. (1989). Changes in regional cerebral oxidative metabolism induced by tactile learning and recognition in man. *Eur. J. Neurosci.*, *1*(1), 3–17.

Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.*, *2*(8), 539–550.

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks. II. Decreases in cerebral cortex. *J. Cogn. Neurosci.*, *9*(5), 648–663.

Shulman, R. G., Hyder, F., & Rothman, D. L. (2001). Cerebral energetics and the glycogen shunt: Neurochemical basis of functional imaging. *Proc. Natl. Acad. Sci. USA*, *98*, 6417–6422.

Shulman, R. G., Rothman, D. L., Behar, K. L., & Hyder, F. (2004). Energetic basis of brain activity: Implications for neuroimaging. *Trends Neurosci.*, *27*, 489–495.

Sibson, N. R., Dhankhar, A., Mason, G. F., Behar, K. L., Rothman, D. L., & Shulman, R. G. (1997). In vivo ^{13}C NMR measurements of cerebral glutamate synthesis as evidence for glutamate-glutamine cycling. *Proc. Natl. Acad. Sci. USA*, *94*, 2699–2704.

Sibson, N. R., Dhankhar, A., Mason, G. F., Rothman, D. L., Behar, K. L., & Shulman, R. G. (1998). Stoichiometric coupling of brain glucose metabolism and glutamatergic neuronal activity. *Proc. Natl. Acad. Sci. USA*, *95*, 316–321.

Siesjö, B. K. (1978). *Brain energy metabolism*. New York: Wiley.

Sokoloff, L., Mangold, R., Wechsler, R., Kennedy, C., & Kety, S. S. (1955). The effect of mental arithmetic on cerebral circulation and metabolism. *J. Clin. Invest.*, *34*, 1101–1108.

Sprague, J. M. (1966). Interaction of cortex and superior colliculus in mediation of visually guided behavior in the cat. *Science*, *153*(743), 1544–1547.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Zempel, J. M., et al. (2007). Intrinsic function architecture in the anesthetized monkey brain. *Nature*, *447*, 83–86.

Vincent, J. L., Snyder, A. Z., Fox, M. D., Shannon, B. J., Andrews, J. R., Raichle, M. E., & Buckner, R. L. (2006). Coherent spontaneous activity identifies a hippocampal-parietal mnemonic network. *J. Neurophysiol.*, *96*, 3517–3531.

Waldvogel, D., Van Gelderen, P., Muellbacher, W., Ziemann, U., Immisch, I., & Hallett, M. (2000). The relative metabolic demand of inhibition and excitation. *Nature*, *406*, 995–998.

Wong-Riley, M. T. (1989). Cytochrome oxidase: An endogenous metabolic marker for neuronal activity. *Trends Neurosci.*, *12*, 94–101.

74 The Neuroeconomics of Simple Goal-Directed Choice (Circa 2008)

ANTONIO RANGEL

ABSTRACT This paper reviews what is known about the computational and neurobiological basis of simple goal-directed choice. Two features define this type of choice. First, individuals make decisions between stimuli that are associated with different outcomes or rewards. Second, the brain solves the decision problem by (1) computing the distribution of outcomes associated with each stimulus, (2) assigning a value to each stimulus equal to the expected reward generated by those outcomes, and (3) selecting the stimulus with the highest computed value. A typical example of simple goal-directed choice is given by the problem of choosing a meal from a buffet table.

Neuroeconomics studies the computational and neurobiological basis of animal and human decision making. Its goal is to understand how the brain solves the multitude of choice problems that organisms face every moment of their existence. One important complication in addressing these problems is that decision-making situations come in many different flavors, and it is likely that the brain uses different computations and systems to solve them. Compare, for example, the problem of a lion chasing a gazelle with the problem of a typical consumer deciding which of two cereal boxes to purchase. Both organisms are engaged in decision making, but their problems are very different. The problem of the lion is to select a direction of movement every instant to increase the probability of catching the gazelle. This entails a simple goal ("catch the gazelle"), but a series of action choices. In contrast, the consumer faces a complicated choice between goals ("which cereal box has the best taste-health-price combination?"), but once that decision has been made, the choice over actions is trivial ("pick the motor plan that grabs the chosen cereal box"). Given this complexity, an important task for neuroeconomics is the construction of a neurally relevant taxonomy of choice tasks that can be used to guide the research and to organize the findings.

Another difficulty in neuroeconomics is that there does not seem to be a simple one-to-one mapping between decision-making situations and the neural processes used to make choices. Instead, a sizable and rapidly growing body of animal and human evidence suggests that there are at least three conceptually and neurally separable behavioral controllers at work in most decision-making situations: a Pavlovian system, a habitual system, and a goal-directed system (Balleine, Daw, & O'Doherty, 2008; Daw, Niv, & Dayan, 2005; Dayan, 2008; Dickinson & Balleine, 2002; Rangel, Camerer, & Montague, 2008). Although this topic is just beginning to be explored, the existing evidence suggests that the relative importance of the systems changes with the details of the decision-making situation. Given these two complications, it is unlikely that we will find a simple neuroeconomic theory of decision making that covers all types of choice situations.

In order to deal with these two complications, research in neuroeconomics typically focuses its attention on a subset of the behavioral controllers and a well-defined subclass of choice problems. In this chapter we review what is known about a class of problems that has received considerable attention in neuroeconomics and behavioral neuroscience: How does the goal-directed system make choices among sets of stimuli associated with different rewards? We refer to this problem as simple goal-directed choice. The decision-making situations of interest resemble the example of the consumer who has to choose one type of cereal among several options. The consumer cares about which choice he makes because the different stimuli are associated with different combinations of outcomes (or rewards). For example, one cereal box might be tastier and cheaper than another. It is important to emphasize that animals also engage in this type of choice. As an example, consider the problem of a rat that has to press a left or a right lever on a Skinner box in order to obtain one of two rewards, or the problem of a hungry lion confronted with several gazelles.

The review has several goals. First, we show that several choice tasks that have been used in the neuroscience and animal learning literatures are special cases of simple goal-directed choice. Second, we provide a mathematical description of the computations that define the goal-directed system. Third, we use the computational framework as a way to organize what is known about the neurobiology of

ANTONIO RANGEL Humanities and Social Sciences, Computational and Neural Systems, California Institute of Technology, Pasadena, California

the goal-directed system and what are some of the most important open questions.

Simple binary stimulus choice: A behavioral paradigm to study simple goal-directed choice

In a simple binary stimulus-choice task, individuals make repeated choices between pairs of stimuli that are presented to them, one on the left and one on the right. Individuals care about their choice because the stimuli are associated with different outcomes (or rewards) that affect their well-being. They indicate their choice by executing one of two different actions associated with each of the stimuli (e.g., a left or a right button push, a left or a right saccade, etc.). The actions are such that the costs and effort required to execute them are as similar as possible.

A typical example of such a task is depicted in figure 74.1*A* (Karjbich, Armel, & Rangel, under review). Individuals are shown pairs of high-resolution pictures of familiar snack-food items in a computer monitor and have to choose which one they would like to consume at the end of the experiment by pressing either a left or a right button. There are 70 different such stimuli that are randomly assigned into pairs in 100 different trials. At the end of the experiment, one of the trials is selected at random, and the subject eats the food depicted in the picture that he chose in that trial. Another example is shown in figure 74.1*B* (Baxter & Murray, 2002). During an initial training phase, monkeys are exposed to 120 objects of different shape and color, two at a time. Importantly, 60 of the objects are associated with a food reward (half with a cherry and half with a peanut) that is placed below the object, whereas the other 60 objects are associated with no such reward. The goal of this phase is for animals to learn to associate the 30 cherry objects with the consumption of a cherry and the 30 peanut objects with the consumption of a peanut. During a second training phase, the animals are repeatedly presented with one cherry object and one peanut object, and are taught to make a choice by lifting only one of the objects and grabbing the reward underneath. The location of the objects is fully randomized. After the animals are fully trained, they are tested in one of three conditions: (1) sessions that are preceded by feeding to satiety with cherries, (2) sessions that are preceded by feeding to satiety with peanuts, and (3) sessions with no prefeeding.

Note some of the central elements of the simple binary stimulus-choice task. First, there are at least 2 stimuli that the subjects choose from, although the set may be much larger. For later reference, let S denote the set of stimuli and s denote a typical element. Second, each stimulus is associated with a probability over outcomes. Let O denote the set of potential outcomes, o denote a typical outcome, and $p(o|s)$ denote the probability that the subject gets outcome o if stimulus s is chosen. The outcomes can be appetitive (e.g., food) or aversive (e.g., a shock). A stimulus might be paired with multiple outcomes. In the simple tasks we have described, the stimulus-outcome associations are degenerate and time-invariant probability distributions, but we need the more general notation to accommodate other tasks of

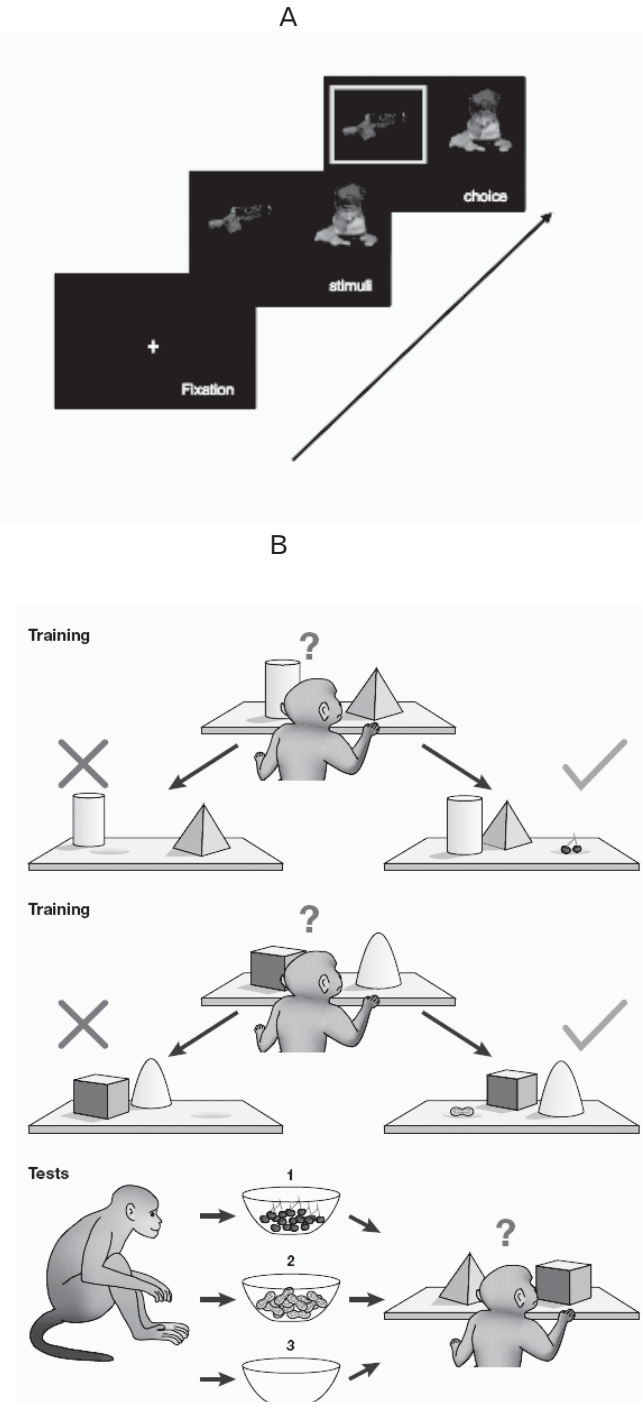


FIGURE 74.1 Examples of simple binary stimulus choice tasks. (*A*) Binary food choice from Karjbich, Armel, and Rangel (under review). (*B*) Devaluation choice task from Izquierdo, Suda, and Murray (2004). (With permission from Baxter & Murray, 2002.) (See color plate 88.)

interest. For example, in reversal-learning tasks the stimulus-outcome associations change with time, and thus we have to write $p^t(o|s)$. Third, the mapping from stimuli to action, denoted by $a^t(s)$, changes from trial to trial. As a result, there is not a fixed mapping between actions and stimuli, or actions and outcomes. This approach captures the fact that in the real world the actions required to obtain a particular stimulus often change over time. Fourth, all the actions required to make or implement the choice entail approximately the same costs to the individual. An example would be a Skinner box with two levers that have equal tension, symmetric location, and so on. The property would be violated if one of the levers is more difficult to pull. Note that in order to keep things simple we make an explicit distinction between the potential aversive outcomes associated with a stimulus (which occur at the time of outcome consumption) and the costs associated with taking the action necessary to obtain the stimulus (which occur at the time of choice).

Another important feature of the binary stimulus-choice task is previous experience consuming all the possible outcomes in all the states of the world that might be induced by the experimenter. To understand why this is important, consider the devaluation experiment depicted in figure 74.1*B*. Here monkeys are asked to make choices among stimuli in three different states of the world: a cherries-satiation condition, a peanut-satiation condition, and a no-satiation condition. To qualify as a simple binary stimulus-choice task, monkeys must have had extensive experience consuming the cherries and peanuts in the three states of the world prior to the actual experiment. As we will see, the choices made by the goal-directed system will depend on how it values the outcomes associated with each stimulus *given the state of the world*. If the subjects have not had experience consuming the outcomes in a particular state, they might need to learn how to evaluate them, a phenomenon that Balleine has called incentive learning (Balleine & Dickinson, 1998). This might result in unstable choices across the experiment. The simple binary stimulus-choice paradigm removes this complication by requiring that subjects have extensive prior experience with all the outcomes in all the relevant states of the world.

The other details of the task are not important and can take many different forms. For example, the stimuli could be pictures on a computer screen, or physical objects with different shape and color, or cards with printed photographs or verbal descriptions of rewards, or even real exposure to the actual outcomes. Subjects might get a reward after every decision, or might get rewarded only for a random subset of the choices that they made at the end of the trial. There are also no constraints on the actions associated with choosing a stimulus, as long as they satisfy the equal-cost property. Thus subjects might indicate their choice through an eye movement and then get the chosen liquid delivered to their mouths, or they may indicate their choice through the act of reaching for one of the stimuli in order to consume it.

The binary stimulus-choice paradigm outlined here covers as special cases several tasks that have been used in the neuroscience and animal learning literatures. First are the type of simple binary choices described in figure 74.1*A* (Kable & Glimcher, 2007; Karjbich et al., under review; Padoa-Schioppa & Assad, 2006, 2008; Tom, Fox, Trepel, & Poldrack, 2007; Wallis & Miller, 2003). Second are the devaluation choice tasks described in figure 74.1*B* (Izquierdo, Suda, & Murray, 2004; Wellman, Gale, & Malkova, 2005). The main differences from the previous set of tasks are that subjects indicate their choice by lifting an object, instead of pressing a button or executing a saccade, and that the value of the outcomes is manipulated by feeding the subject to satiation on some of the foods. Third are reward preference tasks (Izquierdo et al.). The key difference from the previous task is that subjects are exposed to the actual rewards, instead of stimuli associated with them, and they indicate their choice by reaching for the chosen outcome. Fourth are reversal learning tasks (Hampton, Bossaerts, & O'Doherty, 2006). In a typical version of these tasks there are two stimuli and one potential outcome. In every trial the probability of obtaining the outcome is high for one of the stimuli and low for the other, and it evolves over time either through an exogenously specified process or as a function of the history of choices.

It is important to emphasize that, as general as it is, this behavioral paradigm does not cover many decision situations that have also been used in the literature to study the goal-directed system. It rules out the case of multistimulus (nonbinary) choice. It also rules out a popular odor discrimination task from the rat literature (Schoenbaum, Chiba, & Gallagher, 1998) in which rats decide whether or not to drink a liquid from a single location based on the odor that they receive in an odor port (some odors predict rewards like sugar, others punishers like quinine). Note that instead of a choice between stimuli, this task entails choice between motor plans with fixed state-dependent action-outcome relationships. More generally, the paradigm also rules out any tasks in which there is a constant mapping between actions and stimuli or outcomes at each state of the world. It also rules out instrumental paradigms in which animals engage in free rates of responding, even if they have a choice among multiple responses (Balleine & Dickinson, 1998; Dayan & Balleine, 2002). This last class of decision tasks is substantially more complicated as animals need to decide not only what to choose, but also about when to take action. In contrast, in the binary stimulus-choice paradigm the timing of decision making is controlled by the experimenter.

Multiple behavioral controllers: What is goal-directed choice?

As mentioned at the beginning of the chapter, a growing body of evidence suggests that the brain might deploy different behavioral controllers in parallel in many decision-making situations. In this section we provide a brief review of the computational differences among the three main systems that have been identified: a habitual, a goal-directed, and a Pavlovian system. For more detailed reviews see Balleine and colleagues (2008), Dayan (2008), and Rangel and colleagues (2008).

Goal-Directed System The defining feature of the goal-directed system is that it makes choices over stimuli using model-based computations of value. These are carried out in two steps. First, a value is assigned to each stimulus by identifying the distribution of outcomes associated with it and computing the expected values of those outcomes in the current state of the world. Second, the computed values are compared in order to select one of the stimuli.

Note several key properties of this system. First, its goal is to make choices over stimuli, not actions. Second, it assigns value to stimuli by anticipating the outcomes to which they might lead and then computing their expected reward. Third, this computation is based on stimulus-outcome associations and beliefs about the reward that those outcomes are likely to generate in the current state of the world. It follows that the computation is model based (sometimes called forward looking) and is not based on the historical level of payoff generated by the different stimuli. This last property gives enormous flexibility to the system, since it allows it to rapidly update the value that it assigns to stimuli based on either a change on the stimulus-outcome associations or a change in the state of the world that affects their expected value. Fourth, this flexibility comes at the cost of computational complexity. The brain needs to store or compute stimulus-outcome associations and state-dependent value functions, and then needs to carry out expected value computations online.

Habitual System The defining feature of the habit system is that it makes choices over actions based on the historical level of rewards that they have generated. This is also done in two steps. First, a value for each of the available actions is retrieved from memory. Second, the retrieved values are compared in order to select one of the stimuli.

Note several key properties of this system. First, choice is made over actions, not over stimuli. Second, the values used to make choices are retrieved from memory, not computed online. Third, the values assigned to actions depend on the level of rewards that they have generated in the past. Multiple studies have shown that relatively simple reinforcement learning algorithms approximate well the process of value learning for this system (Montague, Dayan, & Sejnowski, 1996; Niv & Montague, 2008; Schultz, Dayan, & Montague, 1997; Sutton & Barto, 1998). Fourth, the fact that the learning can be state sensitive leads to the use of state-dependent action values by the habit system. Fifth, the computations made by the habitual system at the time of choice are simpler than those of the goal-directed system, since values are retrieved from memory instead of computed online. Sixth, this computational simplicity comes at the cost of some behavioral flexibility. Although with enough experience the habitual system is able to make optimal decisions in environments that are sufficiently stable, it cannot do so when the action-outcome contingencies are rapidly changing (as, for example, in the simple experiment described in figure 74.1*A*).

Pavlovian System In contrast to the previous two systems, which are able to assign values to any stimulus or action, the Pavlovian system assigns values to a small set of actions that are evolutionarily appropriate responses to particular environmental stimuli. Typical examples include preparatory behaviors (such as approaching cues that predict the delivery of food) and consummatory responses to a reward (such as pecking at a food magazine).

Although many Pavlovian behaviors are "hardwired" responses to specific predetermined stimuli, with sufficient experience animals can also learn to deploy them in response to other stimuli. For example, rats and pigeons learn to approach lights that predict the delivery of food. At first glance, Pavlovian behaviors look like automatic, stimulus-triggered responses, and not like instances of value-based choice processes. However, since Pavlovian responses can be interrupted, they must be assigned something akin to a "value" so that they can compete with the actions that are favored by the other valuation systems.

The computational and neurobiological basis of the Pavlovian system is much less well understood than that of the habitual and the goal-directed systems. For recent reviews see Dayan and Seymour (2008) and Rangel and colleagues (2008). This lack of understanding is due, in part, to the fact that there might be multiple Pavlovian controllers, some responsible for triggering outcome-specific responses (e.g., pecking at food or licking at water) and others responsible for triggering more general valence-dependent responses (e.g., approaching positive outcomes and withdrawing from negative ones). Nevertheless, since a wide range of human behaviors with important economic consequences might be controlled by the Pavlovian system (from overeating to the harvesting of immediate smaller rewards at the expense of larger delayed rewards), a detailed

understanding of this system is an important open question for neuroeconomics.

Coexisting and Competing Valuation Systems All these behavioral controllers can potentially be active at the same time even in the case of simple binary stimulus choice. Consider, for example, the experiment in figure 74.1*B*. Since some of the stimuli covering the food rewards have been associated with appetitive outcomes, they might trigger Pavlovian approach responses that could influence which of the two objects the monkey lifts first. Similarly, since the execution of the choice entails two constant motor actions (reach for the left object or reach for the right object) and the monkeys receive extensive experience in the task, the habitual system might use historical action values to influence the choice that is made. Finally, the goal-directed system could also bias the monkey's actions by assigning a higher value to the actions associated with the higher-value stimulus. This possibility leads to a very important open question in neuroeconomics about which next to nothing is known: How does the brain assign control to the three different systems? Although some simple computational models have been proposed (Daw et al., 2005; Dayan, Niv, Seymour, & Daw, 2006), to date no experiments have been performed to study how the systems interact and compete at the neural level in simple binary stimulus choice.

In this review we focus on the computations of the goal-directed system during the simple binary stimulus-choice task. We do so not because the effects of the other systems in this type of situations are unimportant, but because much more is known about the role of the goal-directed system. We emphasize, however, that a full understanding of simple stimulus choice will require the study of how the other two systems are deployed in this type of task and of how the allocation of control is resolved.

Computational basis of goal-directed choice in the simple binary stimulus-choice paradigm

In this section we provide a mathematical description of the computations that the goal-directed system needs to make in simple binary stimulus-choice situations.

Representation of the Choice Problem The first problem that the system needs to solve is to identify the parameters of the decision-making problem: What are the potential stimuli that could be chosen? What are the actions required to obtain each stimulus? What external and internal state variables might affect the desirability of the different stimuli and actions? Let e be a summary of the internal and external variables determining the state of the world.

This part of the choice process is often ignored in decision-making models by implicitly assuming that the brain always computes these variables correctly. But given the complexity of the world, it is likely that the brain relies on computational shortcuts. Consider, for example, the problem of a shopper in a modern supermarket aisle that contains thousands of different products. When confronted with such complexity, the brain only evaluates and compares a small subset of the possible items. Since an item is chosen only if it is considered, the representation step has a large impact on the choice that is eventually made. Given the large number of external and internal variables that can impact the choice situation, similar issues are likely to arise in the identification of the relevant states of the world. The algorithms and neural processes at work, as well as the limitations on choice performance to which they lead, are just beginning to be understood (Reutskaja, Pulst-Korenhberg, Nagel, Camerer, & Rangel, under review). Basic open questions include the following: How does the brain determine which actions to assign values to and which actions to ignore? Is there a limit to the number of options that animals can consider at a time? How are internal and external states computed?

Stimulus Valuation As we saw before, the goal-directed system makes choices by assigning values to the different stimuli based on the expected value of the outcomes associated with them. Let $V(s|e)$ denote the value of stimulus s given the state of the world e. In order to compute this value, the system needs two pieces of information: (1) the stimulus-outcome associations, which are summarized by the function $q(o|s)$ specifying the probability that every potential outcome o occurs as a function of the stimulus s, and (2) the value function $v(o|e)$ specifying the value of each outcome given the state of the world.

Note several things about this notation. First, there is a difference between the $p(o|s)$ function that describes the objective mapping between stimulus and outcomes and the $q(o|s)$ function that describes the beliefs of the subject about that relationship. Second, by assumption, the stimulus-outcome associations do not depend on the state of the world, and, for simplicity, we assume that the subjects always know this fact. Third, the value function $v(o|e)$ is the goal-directed system's belief about the reward that it will experience if the outcome occurs, which is a different signal than the level of reward that actually occurs at the time of consumption. Fourth, the value function $v(o|e)$ does not depend on the stimulus. The reason is that there is a conceptual distinction between the (positive or negative) outcomes generated by a stimulus and the costs of taking the action necessary to get that stimulus.

The value assigned to a stimulus is simply the expected value of the outcomes to which it might lead. This is given by

$$V(s|e) = \sum_{o \in O} p(o|s)v(o|e)$$

STIMULUS CHOICE The brain uses the net-value information to make a choice between the stimuli. A sizable amount of behavioral evidence suggests that the maximization process is stochastic and well approximated by a *soft-max* process in which the probability of choosing stimulus s is given by

$$\frac{e^{\tau V(s|e)}}{\sum_{t \in S} e^{\tau V(t|e)}}$$

where τ is a coefficient measuring the sensitivity of the choices to the stimulus values (when $\tau = 0$ each alternative is chosen with equal probability regardless of the values, and for sufficiently large τ almost all of the probability falls on the item with the highest value).

The soft-max model is a reduced-form model of limited use for neuroeconomics, since it describes how the probability of making a choice changes with the net values, but not how the choice is actually made. A large research effort is devoted to this problem (for recent reviews see Bogacz, 2007; Busemeyer & Johnson, 2004; Ditterich, 2006; Gold & Shadlen, 2007; Rangel, 2008). Most of the models that have been proposed are versions of a race-to-barrier diffusion process. A simple version of the model for the case of two alternatives is depicted in figure 74.2*A*. The model has several components. First, there are circuits that compute the value of each of the items. The value assigned to the items is assumed to fluctuate stochastically from instant to instant. Every instant, the two value signals are subtracted to produce a relative-value signal that is then fed to an integrator circuit that computes the value of item 1 minus the value of item 2, thus keeping track of the accumulated relative signal. A decision is made when this relative-value signal becomes sufficiently large ("choose item 1") or sufficiently negative ("choose item 2").

This class of models has several attractive features. First, they predict a logistic choice function similar to the one generated by the soft-max model. Second, they predict that the time required to make a choice should be larger when items have similar values than when the values are far apart.

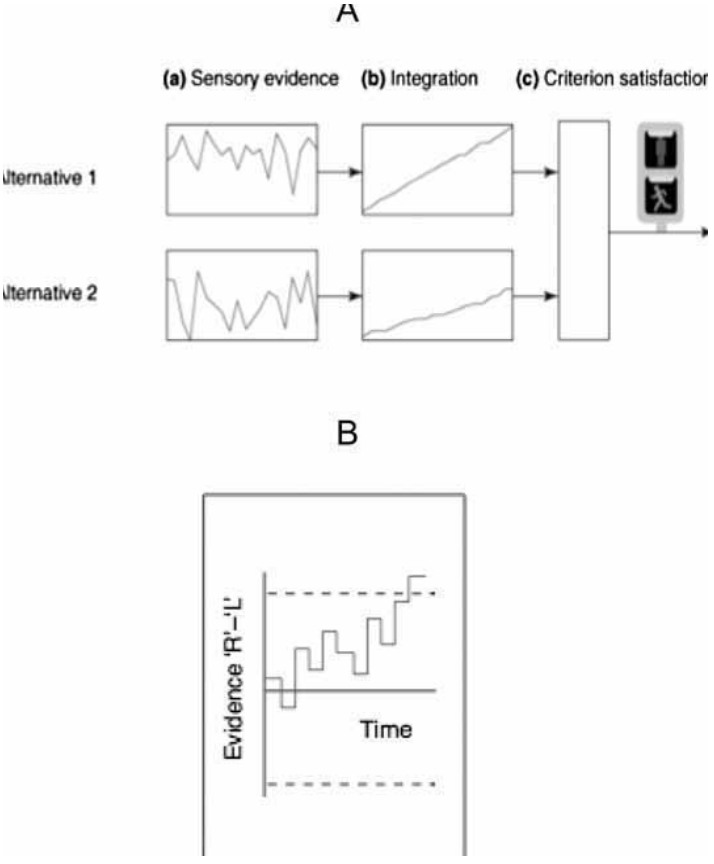


FIGURE 74.2 Models of the value comparison process. (*A*) Illustration of the main components of the race-to-barrier models. (Adapted with permission from Bogacz, 2007.) (*B*) A typical run of the random walk model. The step function represents the accumulated relative value of the "right" target. The process starts at a middle point and stops the first time this variable crosses one of the thresholds (depicted by the bracketed horizontal lines). "Right" is chosen when it crosses the upper threshold; "left" is chosen when it crosses the lower one. Time advances in discrete steps. The size of every step is given by a Gaussian distribution with a mean that is proportional to the true direction of motion. This noise is meant to capture the variability in the valuation processes. (See color plate 89.)

Both predictions are consistent with a large body of behavioral data. Finally, the model makes useful predictions about which kind of computations should be implemented by the brain: there should be circuits computing the value of each stimulus, circuits computing relative values, an integrator circuit, and a circuit that triggers a choice when a barrier is crossed.

These models assume that a choice is first made over stimuli and that the choice is then implemented by deploying the action that leads to that stimulus. We refer to these types of models as *stimulus-based choice*. Another a priori equally plausible theory (Glimcher, Dorris, & Bayer, 2005) specifies that the brain uses the stimulus values to assign a value to every feasible action, and that it then makes the choice through a process of competition over action plans. We refer to this possibility as *action-based choice*. It is difficult to compare these two views on theoretical grounds, since the race-to-barrier models apply well to both types of choice. Thus novel experiments are needed to address this issue. The question is important because the neural systems involved in making the choice are likely to be different in the case of stimulus- and action-based choice.

Learning In some versions of the simple binary stimulus-choice paradigm, subjects receive an outcome after every choice. This provides them with feedback that can be used to update their estimate of the stimulus-outcome associations. Here we propose a simple algorithm that subjects can use to carry out this type of learning.

We assume that the experimental task is structured as follows. Every experimental trial t begins with the revelation of the current state of the world (e^t). A stimulus s^t is then chosen that leads to the set of outcomes O^t and a level of reward r^t.

We assume that learning takes place in two stages. In the first stage a prediction error is computed for every possible outcome in the set O. These prediction errors are given by

$$\delta^t(o) = I_O - q^t\left(o \middle| s^t\right)$$

where I_0 is an indicator function taking a value of 1 if the outcome in question occurs and a value of zero otherwise. Note that positive prediction errors measure the degree to which the occurrence of an outcome was surprising, and negative prediction errors measure the extent to which the nonoccurrence of the other outcomes was surprising. In the second stage the prediction errors are used to update the state-outcome probability function for that stimulus by

$$q^{t+1}\left(o \middle| s^t\right) = q^t\left(o \middle| s^t\right) + \lambda\delta^t(o)$$

where λ is a learning rate between 0 and 1 that affects the speed of learning.

Note that this formulation assumes that only the beliefs for the stimulus that was chosen are updated. This approach assumes a very strong form of discrete learning, an assumption which is plausible in environments where there are a small number of highly dissimilar stimuli but not in domains in which "similar" stimuli have "similar" stimulus-outcome associations. In the later case, the outcome in one state can provide information about the stimulus-outcome associations for other states. The extent to which the goal-directed system engages in this type of generalization is largely unknown.

Neurobiological basis of goal-directed choice in the simple binary stimulus-choice paradigm

In this section we review some of what is known about how the brain implements the computations described in the previous section and highlight some important open questions. For alternative recent reviews see Balleine and colleagues (2008), Rangel (2008), Rangel and colleagues (2008), and Wallis (2007).

Representation Unfortunately, next to nothing is known about this important step in the decision-making process. Open questions of particular interest include the following. How does the brain know when to activate the goal-directed evaluation and comparison circuitry? How does the brain decide which stimuli to evaluate at any given moment? Which aspects of the state of the world are measured, and how are they encoded by the goal-directed evaluation circuitry? The first question is important because organisms are exposed to potential choice stimuli continuously, but the goal-directed choice might only engage in the process of choice sporadically. The second question is important because often there are many potential stimuli and the system might not have the capacity to evaluate all of them fully. Think, for example, of a consumer in a modern supermarket aisle.

Stimulus Valuation Several papers have found neural correlates of the stimulus-value signal ($V(s|e)$). Plassmann, O'Doherty, and Rangel (2007) investigated the neural correlates of stimulus valuation by the goal-directed system in humans using fMRI. They showed pictures of desirable snacks to hungry subjects who had to place bids for the right to eat them at the end of the experiment. The size of the bids was a measure of the value assigned by the brain to each stimulus at the time of choice and positively correlated with BOLD activity in the mOFC and the dorsolateral prefrontal cortex (DLPFC). (For related fMRI findings see Arana et al., 2003; Erk, Spitzer, Wunderlich, Galley, & Walter, 2002; Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008; Paulus & Frank, 2003; Tom et al., 2007; Valentin, Dickinson, & O'Doherty, 2007).

A related study used single-unit electrophysiology in nonhuman primates to look for activity in the orbitofrontal cortex that correlates with stimulus values (Padoa-Schioppa

& Assad, 2006, 2008). Every trial, thirsty animals were given a choice between two stimuli associated with small magnitudes of two different juices. After a period of deliberation, the animals indicated their choice with a left- or right-eye movement. The action associated with each stimulus varied from trial to trial. The authors estimated a logistic-choice model to compute a measure of value for each juice-amount combination that was then correlated with the neural signals. They found a large population of neurons encoding the value of the stimulus associated with each juice independently of the action that it took to get it. They did not find an equivalent population encoding the value of the actions. A closely related study recorded simultaneously from monkey OFC and DLPFC and found neurons encoding for the value of stimuli in both areas, although the value signal arose in DLPFC with a delay of approximately 100 ms (Wallis & Miller, 2003).

The previous studies looked for stimulus-value signals in the case in which animals made choices between appetitive items. An important question is whether the brain uses the same networks to evaluate stimuli associated with aversive items (e.g., choosing which of two undesirable risks to take). Plassmann, O'Doherty, and Rangel (under review) used an experimental design similar to the one we have described to study this question. Subjects were shown pictures of undesirable food items (e.g., canned vegetables) and had to bid to avoid having to eat them. The bids were a measure of the extent to which they disliked the foods. Interestingly, no areas exhibited a positive and significant correlation with this measure of stimulus value. Instead, the study found that activity in the mOFC and the DLPFC were negatively correlated with the bids. This finding suggests that these two structures play a role in the valuation of both appetitive and aversive items, in the appetitive case through increased activity and in the aversive case through decreased activity.

Interestingly, given that the $V(s|e)$ function is a forecast of the actual value of consuming the objects associated with the stimuli, activity in the OFC has also been shown to be correlated with the value of expected outcomes in the absence of choice. For example, Gottfried, O'Doherty, and Dolan (2003) presented subjects with visual stimuli that were paired with different odors and used a devaluation procedure to manipulate the value of some of the odors. Using human fMRI, they found that activity in amygdala and OFC was consistent with the encoding of the expected odor value at the time of cue presentation (prior to the actual odor delivery). (For related human fMRI studies see Gottfried, O'Doherty, & Dolan, 2002; Nobre, Coull, Frith, & Mesulam, 1999; O'Doherty, Deichmann, Critchley, & Dolan, 2002). These findings, together with the ones for goal-directed choice described previously, suggest that the OFC might be involved in the computation of different types of value signals at different stages of the choice process and in different types of tasks.

In all the previous experiments, there were no costs associated with choosing an item. Hare and colleagues (2008) studied a simple choice paradigm in which subjects had to make a decision about whether or not to buy a food snack at a given price. In this case, acquiring the stimulus entailed a cost equal to a loss of money given by the price. Consistent with the studies described before, they found that the value of the foods correlated with activity in the medial OFC, but that the price was not encoded in this area. Instead, a "consumer surplus" signal, equal to the value of the item minus its price, was found in the central OFC. These results suggest that the medial OFC might be involved in the encoding of stimulus value but is not responsive to the costs of acquiring the item.

A difficulty in identifying areas where stimulus values might be encoded is that these signals are most likely positively correlated with other signals that are not part of the goal-directed-system valuation process. Consider several examples that have caused some confusion in the literature. First, exposure to stimuli with very positive or very negative stimulus values might induce an increase in arousal in systems associated with motor preparation. If the experimental condition only includes appetitive items, the arousal and stimulus-value signals will be perfectly correlated, and thus one might misattribute one type of signal for the other. As proposed by Roesch and Olson (2004), one way of dissociating the two signals is to include both appetitive and aversive items in the experiment: neural value signals increase linearly with stimulus value, whereas arousal signals are correlated with the absolute value of the stimulus value. Using this logic in a monkey electrophysiology experiment, Roesch and Olson found that activity in OFC reflected the stimulus value, whereas activity in premotor cortex reflected an arousal-type variable. Second, similar to the case of arousal, exposure to stimuli with very positive or very negative stimulus values might induce an overall increase in attention. Third, in many choice paradigms, goal values and reward prediction errors are positively correlated (even if the design includes both appetitive and aversive items). Hare and colleagues (2008) show that prediction errors and stimulus values can be dissociated by introducing a random monetary prize in every trial that is independent of the choices made by the subjects. Using this experimental trick, they found that BOLD activity in the medial OFC, but not ventral striatum, was correlated with the stimulus values, whereas activity in the ventral striatum was most consistent with the prediction error signal.

Why is medial OFC involved in the computation of stimulus values? Some authors have argued that this area of the prefrontal cortex might be in a unique position to integrate information about stimuli and states of the world into a value

(Schoenbaum, Roesch, & Stalnaker, 2006; Wallis, 2007). This favorable position is due to its multiple connections with limbic areas such as the thalamus, amygdala, and striatum (Carmichael & Price, 1996; Ongur & Price, 2000).

So far we have focused on the neural basis of the stimulus-value signal. This signal represents only the output of the valuation process. As described in the previous section, these values are constructed by either retrieving or computing stimulus-outcome associations [the $p(o|s)$ functions], by retrieving or computing the value associated with each of those outcomes in the current state of the world [the $v(o|e)$ function], and by integrating them into an expected value signal [the $V(s|e)$ stimulus value]. This analysis gives rise to the following important questions: How and where are the stimulus-outcome associations represented? How and where is the $v(o|e)$ valuation function represented, and how does the state of the world modulate its value? How and where are the two of them integrated into the stimulus value signal? The answers to these questions are largely unknown and constitute one of the most important open problems in neuroeconomics.

Stimulus Choice Although several proposals have been made about how the brain compares options in simple stimulus-choice situations (Glimcher et al., 2005; Wallis, 2007), next to nothing is known about this is actually done. Understanding how the goal-directed systems compare the stimulus values to make a choice is another important open problem in neuroeconomics. Other open questions include the following: Does the brain make choices by implementing a race-to-barrier model? If so, is the choice made over actions or stimuli? How are the barriers chosen and implemented? How does the slope of the integrators relate to the strength of the stimulus-value signal encoded in medial OFC? Are there other inputs to the comparison process besides the medial OFC signal? How and where does the brain incorporate information about the cost of acquiring the different stimuli? How does the system go from stimulus choices to motor responses?

Learning There is a large literature in neuroeconomics showing that reward prediction errors are encoded in the ventral striatum in the context of Pavlovian (nonchoice) and habitual choice paradigms (for a comprehensive review see Niv & Montague, 2008). Unfortunately, this literature is not very informative about the learning that takes place in the goal-directed system during the simple stimulus-choice task. The reason is that the prediction errors required here measure how surprising the occurrence of individual outcomes is, as opposed to prediction errors of reward that measure the amount of unexpected reward received at the time of consumption. These are two very different types of learning and are likely to be implemented by different networks. Understanding the computational and neurobiological basis of how the goal-directed system learns stimulus-outcome associations is another important open question for neuroeconomics.

Conclusions

The goal-directed system provides organisms with a flexible and adaptive tool to make decisions. This is based on its ability to assign values to stimuli based on beliefs about the outcomes that they are likely to generate and the value of those outcomes in the current state of the world. The system might be particularly powerful through its interactions with other higher cognitive processes that might allow it to use analytical and memory processes to improve its characterization of the stimulus-outcome associations. Given that the system is thought to play a large role in human decision making, understanding its computational and neurobiological basis is central to understanding the essence of human nature.

This review has emphasized the use of simple mathematical models to describe the computations that the goal-directed system needs to carry out in order to make a choice. These models are useful because they lay down precise descriptions of the computational nature of the problem ("what needs to be encoded") and guide the search for the neural instantiation of the process at work. We believe that the use of these types of models is critical to the rapid advancement of the field.

REFERENCES

Arana, F. S., Parkinson, J. A., Hinton, E., Holland, A. J., Owen, A. M., & Roberts, A. C. (2003). Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. *J. Neurosci.*, *23*(29), 9632–9638.

Balleine, B. W., Daw, N., & O'Doherty, J. (2008). Multiple forms of value learning and the function of dopamine. In P. W. Glimcher, E. Fehr, C. Camerer, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision-making and the brain*. New York: Elsevier.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4–5), 407–419.

Baxter, M. G., & Murray, E. A. (2002). The amygdala and reward. *Nat. Rev. Neurosci.*, *3*(7), 563–573.

Bogacz, R. (2007). Optimal decision-making theories: Linking neurobiology with behaviour. *Trends Cogn. Sci.*, *11*(3), 118–125.

Busemeyer, J. R., & Johnson, J. G. (2004). Computational models of decision making. In D. Koehler & N. Narvey (Eds.), *Handbook of judgment and decision making* (pp. 133–154). New York: Blackwell.

Carmichael, S. T., & Price, J. L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *J. Comp. Neurol.*, *371*, 179–207.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, *8*(12), 1704–1711.

Dayan, P. (2008). The role of value systems in decision making. In C. Engel & W. Singer (Eds.), *Better than conscious? Implications for performance and institutional analysis*. Cambridge, MA: MIT Press.

Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, *36*(2), 285–298.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Netw.*, *19*(8), 1153–1160.

Dayan, P., & Seymour, B. (2008). Values and Actions in Aversion. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain*. New York: Elsevier.

Dickison, A., & Balleine, B. W. (2002). The role of learning in the operation of motivational systems. In C. Gallistel (Ed.), *Learning, motivation and emotion*, vol. 3 of *Stevens' handbook of experimental psychology* (3rd ed., pp. 497–533). New York: John Wiley & Sons.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Net.*, *19*(8), 981–1012.

Erk, S., Spitzer, M., Wunderlich, A. P., Galley, L., & Walter, H. (2002). Cultural objects modulate reward circuitry. *NeuroReport*, *13*(18), 2499–2503.

Glimcher, P. W., Dorris, M. C., & Bayer, H. M. (2005). Physiological utility theory and the neuroeconomics of choice. *Games Econ. Behav.*, *52*, 213–256.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*, 535–574.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J. Neurosci.*, *22*(24), 10829–10837.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, *301*(5636), 1104–1107.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.*, *26*(32), 8360–8367.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.*, *28*(22), 5623–5630.

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.*, *24*(34), 7540–7548.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.*, *10*(12), 1625–1633.

Karjbich, I., Armel, C., & Rangel, A. (under review). The role of visual attention in the computation and comparison of economic values.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.*, *16*(5), 1936–1947.

Niv, Y., & Montague, P. R. (2008). Theoretical and empirical studies of learning. In P. W. Glimcher, E. Fehr, C. Camerer, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain*. New York: Elsevier.

Nobre, A. C., Coull, J. T., Frith, C. D., & Mesulam, M. M. (1999). Orbitofrontal cortex is activated during breaches of expectation in tasks of visual attention. *Nat. Neurosci.*, *2*(1), 11–12.

O'Doherty, J. P., Deichmann, R., Critchley, H., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron*, *33*, 815–826.

Ongur, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex*, *10*(3), 206–219.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226.

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.*, *11*(1), 95–102.

Paulus, M. P., & Frank, L. R. (2003). Ventromedial prefrontal cortex activation is critical for preference judgments. *NeuroReport*, *14*(10), 1311–1315.

Plassmann, H., O'Doherty, J., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.*, *27*(37), 9984–9988.

Plassmann, H., O'Doherty, J., & Rangel, A. (under review). Aversive goal values are negatively encoded in the medial orbitofrontal cortex at the time of decision making.

Rangel, A. (2008). The computation and comparison of value in goal-directed choice. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain*. New York: Elsevier.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.*, *9*(7), 545–556.

Reutskaja, E., Pulst-Korenhberg, J., Nagel, R. M., Camerer, C. F., & Rangel, A. (under review). The brain uses a random search and pairwise comparison algorithm to make multi-item choices under time pressure.

Roesch, M. R., & Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science*, *304*(5668), 307–310.

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat. Neurosci.*, *1*(2), 155–159.

Schoenbaum, G., Roesch, M. R., & Stalnaker, T. A. (2006). Orbitofrontal cortex, decision-making and drug addiction. *Trends Neurosci.*, *29*(2), 116–124.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*, 515–518.

Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.*, *27*(15), 4019–4026.

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.*, *30*, 31–56.

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.*, *18*(7), 2069–2081.

Wellman, L. L., Gale, K., & Malkova, L. (2005). $GABA_A$-mediated inhibition of basolateral amygdala blocks reward devaluation in macaques. *J. Neurosci.*, *25*(18), 4577–4586.

75 Neuroeconomics and the Study of Valuation

PAUL W. GLIMCHER

ABSTRACT Just over a decade ago neurobiologists knew almost nothing about the neural mechanisms of voluntary choice. In contrast, economists and psychologists working at that time had well-developed frameworks for describing the many hidden processes that must underlie choice, but these frameworks had very little impact in neurobiological circles. The last decade, however, has seen a revolution in the neurobiological understanding of choice that has been driven by an integration of economics and psychology into mainstream neuroscience. Today, the basic outlines of the primate system for decision making are emerging from studies on humans and monkeys that rely on techniques ranging from single-neuron electrophysiology to functional magnetic resonance imaging (fMRI). Indeed, since the last edition of this book was published a new field for the study of decision making has emerged, *neuroeconomics*, and an edited volume has been published that surveys the field (Glimcher, Camerer, Fehr, & Poldrack, 2008). This chapter provides an outline of the primate mechanism for choice as we understand it today. In broad strokes, we now believe that choice involves a two-stage neural process. The first stage, largely resident in the frontal cortex and the basal ganglia, learns and represents the value of our actions. The second stage, largely resident in a frontoparietal network, selects the option that has the highest subjective value from among the options before us at any moment in time.

Introduction

Our existing data now suggest that when we make a choice we employ a two-step neurobiological process with some remarkable similarities to both psychological and economic process models of decision making. The first step in the neurobiological processes that guide decision making places idiosyncratic valuations on the options before a chooser. These valuations involve the activation of many frontocortical and basal ganglia circuits. The second step chooses, based on those valuations, a single action for execution. Although less well understood than the valuation processes, these choice processes involve both frontal and parietal circuits. What follows is an overview of the valuation and choice mechanisms as they are understood today. Without a doubt, this understanding is fragmentary, and some of the conclusions drawn here will be somewhat controversial, but the presentation captures the state of the field today and suggests just how much has been accomplished since the third edition of this volume was published only five years ago.

PAUL W. GLIMCHER Professor of Neural Science, Economics, and Psychology; Director, Center for Neuroeconomics; New York University, New York, New York

The two-stage model

The neurobiological evidence for a two-stage model emerged initially from studies of decision making in awake behaving monkeys conducted throughout the 1990s by two groups of researchers. The first of these groups was concerned with understanding how animals engaged in traditional psychophysical tasks that required the evaluation of visual stimuli reached a perceptual decision (Newsome, Britten, Salzman, & Movshon, 1990; Gold & Shadlen, 2007). The second emerged from the study of movement control and was concerned with understanding how changes in the magnitude or probability of reward influenced decision making (Platt & Glimcher, 1999; Glimcher, 2002). Both groups converged, however, to the view that neurons of the posterior parietal cortex participated in the actual process of deciding (selecting one action from a finite set of alternatives) and that these neurons received inputs that encoded something about the magnitude or likelihood of future rewards (associated with each of those alternatives) that originated from signals generated elsewhere in the brain.

In Platt and Glimcher's (1999) study, the authors recorded from neurons in the posterior parietal cortex while thirsty monkeys participated in a simple forced-choice task. In that task, monkeys fixated a central yellow target while two eccentric visual stimuli (one red and one green) were coilluminated (figure 75.1). One of those targets was located within the response field of a parietal neuron under study. After a brief delay, the central target then switched color to either red or green, indicating which of the two eccentric stimuli the animal should fixate in order to receive a reward. What the authors varied, across blocks of about 100 trials, was either the magnitude of reward associated with each of the targets or the probability that the fixation target would turn red. They found that, *immediately after target onset*, if the magnitude of reward associated with the target inside the response field was increased, the firing rates of neurons encoding an eye movement to that target increased at the very beginning of

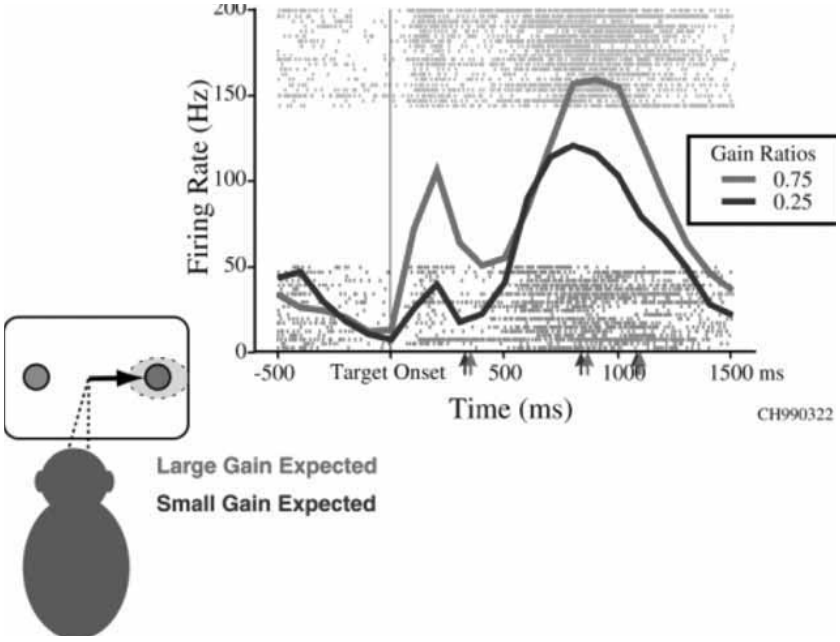


FIGURE 75.1 Same movement, different values. (From Glimcher, 2003.) (See color plate 90.)

the trial. They also showed that if the probability that an eye movement toward the response field would be reinforced was high, the units responded more strongly than if a movement toward that target was unlikely to yield a reward. *Immediately before eye movement onset*, however, the neuronal firing rate indicated whether or not the animal had chosen to produce the saccade encoded by that neuron.

In interpreting this result, they noted that all economic theories of choice predict that valuation should always be influenced by both the probability and the magnitude of reward. This theory suggested that the early activity observed in these neurons might well encode the subjective value of the eye movements to the monkeys. The late activity, in contrast, appeared to encode choice, the output of an operation performed on the set of eye movements available to the animal. The suggestion, then, was that the inputs to these parietal circuits might well encode an idiosyncratic subjective valuation of the kind described by economic theories of choice and that parietal (and related extraparietal) circuits might use these valuation inputs as part of a winner-take-all computation to choose actions for execution.

At the same time that these studies were being conducted, a number of lines of evidence began to suggest that portions of the striatum and the frontal cortex both learn and represent the values of goods and actions—a finding suggesting that these areas might serve as the source of the valuation signals identified in parietal cortex. The critical first step toward this realization was the identification of reinforcement learning mechanisms in the forebrain, and it is an understanding of these learning mechanisms that has paved the way toward a broader understanding of valuation. In the early 1990s, Wolfram Schultz and his colleagues (e.g., Romo & Schultz, 1990; Schultz & Romo, 1990; Schultz, Apicella, & Ljungberg, 1993) demonstrated that midbrain dopaminergic neurons encode a *reward prediction error*. Montague, Dayan, and Sejnowski (1996) provided the next step when they recognized that this class of signal could be employed to construct a mechanism that learns, through trial and error, the values of actions or objects that could be used to guide choice. What followed were 10 years of work that established the existence of at least three interrelated subsystems in these brain areas that employ distinct mechanisms for learning and representing value and that interact to produce the valuations that guide choice (Dayan & Balleine, 2002; Balleine, Daw, & O'Doherty, 2008; Niv & Montague, 2008).

In a similar way, studies of the movement control systems of the brain strengthened the conviction of many that a discrete choice mechanism used these valuation signals to select and execute actions. Our current evidence indicates that the choice system involves large portions of the parietal cortex, among other areas. These parietal areas receive both direct and indirect projections from the valuation areas and project directly to the movement control areas. One issue that remains unclear, however, is how much of the frontal cortex and basal ganglia participate directly in the choice process with these parietal areas. We now know that specific neurons in the orbitofrontal cortex (Padoa-Schioppa & Assad, 2006, 2008) and the dorsal striatum (Samejima, Ueda, Doya, & Kimura, 2005; Lau & Glimcher, 2008) of the monkey also represent goods and actions that have been chosen before these choices are executed, but whether these neurons participate directly in choice is not known at this time.

This then, is a minimal working outline of the primate choice system: A valuation system that learns through repeated sampling of the environment and stores the values of actions and/or goods; a choice system that uses these values to select an action for execution; and a motor control system that executes the physical responses dictated by the choice. Of course, future experiments will enrich this description; for example, it may well be the case that perceptual systems influence the valuation systems in ways that we are just beginning to understand, but these seem to be the fundamental components of the primate architecture for choice as we understand it today.

The basic structure of the valuation system

The critical breakthrough that allowed modern studies of valuation to crystallize were insights into the function of the midbrain dopaminergic pathways. In 1993, Schultz and colleagues measured the spiking activity of single dopamine neurons while monkeys passively received rewards during a classical conditioning task. They found that *unconditioned* rewards produced a strong response in these neurons while conditioned rewards did not. This was an important finding because it revealed that the activity of dopamine neurons *could not* simply code hedonic experience but rather appeared to encode something more closely related to learning itself. This revelation led Montague, Dayan, and Sejnowski (1996) to propose that dopamine neurons encoded the difference between expected and obtained rewards: the *reward prediction error* of reinforcement learning theory. The critical idea that emerged over the next several years was that dopamine spike rates communicated to frontocortical and striatal circuits the degree to which rewards actually obtained by the subject matched previously learned predictions of reward magnitude. This explained why dopamine neurons responded to unconditioned rewards (which the animals did not expect) while remaining silent when animals received conditioned rewards (which the animals expected).

More formally, what these studies suggested was that dopamine neurons coded a term from reinforcement learning theory that had been previously developed within psychological circles. In 1972, Rescorla and Wagner had proposed that the associative strength between a stimulus and a reward during classical conditioning could be described by the rule

$$AssStr_{new} = AssStr_{old} + \alpha(Reward - AssStr_{old})$$

where *AssStr*, or "associative strength," is thus incremented (or decremented) by the difference between the reward obtained and the reward expected (the *old* associative strength) until the prediction matches the experience and learning is thus complete. In this formulation, α is a number between 0 and 1 that controls how gradually learning shifts the prediction across trials from old values to new values. (In fairness, this is not exactly the form in which Rescorla and Wagner made their proposal. Their proposal employed an additional term associated with stimulus salience and also made predictions about how two stimuli competed to predict a single reward. The form shown here is much closer to a model originally proposed in 1951 by Bush and Mosteller that served as the basis of Rescorla and Wagner's later model.)

Subsequent studies of the dopamine neurons and many of their targets have largely validated this early conclusion of Montague's and extended these insights into the domain of operant conditioning in animals. In 1992, Schultz and his colleagues (Ljungberg, Apicella, & Schultz, 1992) showed, for example, that even in a classical conditioning task dopamine neurons encoded a signal that closely paralleled the reward prediction error term of theory. Subsequent studies using more sophisticated computational methods (Bayer & Glimcher, 2005; Morris, Nevet, Arkadir, Vaadia, & Bergman, 2006) have also validated this hypothesis. Together, these data demonstrate unequivocally that dopamine neurons carry a signal to the striatum and frontal cortices that is sufficient to account for how animals learn the values of their actions, clear evidence that a valuation signal of some kind could be constructed and stored in these areas (or their targets) within the brains of monkeys.

Fortunately, there is also clear evidence that these dopaminergic neurons behave in the same manner in humans as they do in monkeys. Like other mammals, humans find dopaminergic drugs reinforcing. Like other mammals, humans have these same dopaminergic pathways. Like other mammals, dopaminergic drugs can be shown to bind to receptors in the terminal fields of these neurons. But the best evidence for the notion that a circumscribed learning-based valuation system associated with dopamine occurs in humans comes from fMRI studies of humans engaged in learning about rewards. In 2002, two groups (O'Doherty, Deichmann, Critchley, & Dolan, 2002; Pagnoni, Zink, Montague, & Berns, 2002) demonstrated simultaneously that activity in the dopaminergic terminal fields of the striatum and the frontal cortex during both gustatory and monetary reward tasks behaved exactly as predicted. This result indicated that there existed dopaminergic signals appropriate for value learning in humans.

Contemporary studies of these neurons continue to extend and refine these conclusions in important ways. We now have reason to believe that the actual algorithm computed by the dopamine neurons or their sources is a more refined version of the Rescorla and Wagner model known as the *temporal difference model* (Sutton & Barto, 1998). This model explains not just how expected rewards are encoded, but how a dopamine-based system could develop associations between stimuli and rewards that are separated in time. The temporal difference model, like the dopamine neurons, is able to connect the ringing of a bell with a food reward

that follows it seconds later (Schultz et al., 1993). This is an important advance, but one that lies beyond the scope of this brief review.

What remains, then, is to understand where and how these dopamine activations are used to mechanistically compute and store the values of actions. Two lines of evidence contribute to our contemporary understanding of these issues: neuronal recording studies in animals and fMRI studies in humans. The recording studies in animals have now established that the basal ganglia (and in particular the striatum) contain essentially all of the computational elements required for the execution of reinforcement learning (or more precisely *temporal difference* learning) algorithms. There are, for example, neurons within the basal ganglia that encode the magnitude of reward that an animal expects to receive for producing a particular behavioral action (Hikosaka, Takikawa, & Kawagoe, 2000; Samejima et al., 2005; Lau & Glimcher, 2008), neurons that encode the actions that have just been executed (Samejima et al.; Lau & Glimcher, 2007), and neurons with firing rates dependent on the current state of the environment (Hikosaka, 2007), among other things. These neurons are located in the striatum and project out of the basal ganglia largely through the ventrolateral nucleus of the thalamus, which projects in turn back to the frontal cortex. Single-unit recording studies in the frontal cortex have also demonstrated the existence of neurons that encode values, but this time the values of goods, not of actions (Padoa-Schioppa & Assad, 2006, 2008). Functional MRI studies in humans tell a similar story (Knutson, Westdorp, Kaiser, & Hommer, 2000; Knutson, Taylor, Kaufman, Peterson, & Glover, 2005; O'Doherty et al., 2002; O'Doherty, Buchanan, Seymour, & Dolan, 2006), suggesting that frontal and basal ganglia circuits form the core of the human mechanism for learning and representing value.

There is, however, evidence for other learning mechanisms in these same structures that interact with this well-studied Rescorla-Wagner-style learning mechanism. The details of these other learning systems are still being worked out, but in essence these studies suggest that a set of mechanisms, most if not all interacting with dopamine, provide tools for learning and representing value in the frontal cortex and the basal ganglia (Balleine et al., 2008).

For neuroeconomists, these studies constitute overwhelming evidence that a value system exists and can be functionally localized. Where, then, is the final point of convergence at which these values that guide choice, likely computed by several interaction neural circuits organized around the frontal cortex and the basal ganglia to the choice system, are acted on by the choice system that guides action?

One way to begin to answer this question is to look at the existing fMRI data and to ask, Are there a small number of areas that are actively correlated with subjective value under essentially all reward and choice conditions that have ever been studied? Perhaps surprisingly, the answer to this question seems to be yes. The ventral striatum and the medial prefrontal cortex show up in dozens of studies under essentially all choice conditions as coding something like the values we infer humans and animals place on their own actions.

Activity in the ventral striatum has been shown to be correlated with both rewards and punishments (Delgado, Nystrom, Fissell, Noll, & Fiez, 2000), the magnitude of cumulative rewards (Elliot, Friston, & Dolan, 2000), the anticipation of reward (Knutson et al., 2000; Knutson, Fong, Bennett, Adams, & Hommer, 2003), the expectation of monetary reward (Breiter, Aharon, Kahneman, Dale, & Shizgal, 2001), the expectation of primary rewards (O'Doherty et al., 2002), the receipt of monetary rewards (Elliott, Newman, Longe, & Deakin, 2003), monetary expected values (Knutson, 2005), behavioral preference rankings among rewards (O'Doherty et al., 2006), potential gain magnitude and loss magnitude as scaled by subject-specific levels of loss aversion (Tom, Fox, Trepel, & Poldrack, 2007), and discounted reward value at delays ranging from minutes to six months (Kable & Glimcher, 2007). Single-unit recording studies in the dorsal striata of monkeys, both in the caudate (Lau & Glimcher, 2007) and the putamen (Samejima et al., 2005), tell a similar story. Neurons that clearly code action values have been identified in these areas. All these data suggest that whenever rewards are received or preferences are expressed, activity in the ventral striatum encodes the magnitudes of those rewards or preferences.

A similar correlation seems to hold in the medial prefrontal cortex. Activity in this area has been shown to be correlated with monetary reward magnitude (Knutson et al., 2001, 2003), preference ordering among primary rewards (McClure, Li, et al., 2004), the expected value of a lottery (Knutson et al., 2005), the subject-specific valuation of gains and losses (Tom et al., 2007), subject-specific discounted reward value (Kable & Glimcher, 2007), and willingness to pay (Plassmann, O'Doherty, & Rangel, 2007). Activity in this area appears to be correlated with valuation under all these conditions. These data have led to the proposal that mean activity in the medial prefrontal cortex and the ventral striatum serves as a final common path for encoding the values of actions (Glimcher, Dorris, & Bayer, 2005).

It should be noted, however, that this conclusion remains somewhat controversial. An alternative hypothesis active in the literature proposes that the valuations we infer from behavior reflect the interaction of two or more largely independent neural systems that compete to govern behavior, the so-called *multiple-self* models. These models typically propose the existence of two largely independent decision-making systems: one associated with so-called *limbic* areas of the brain and the other with so-called *rational* areas of the brain. While tremendously interesting from an economic point of view,

these models are, for the most part, at variance with most of the existing corpus of neurobiological data. Still, it is germane to ask whether the existing evidence supports a two-agent model of decision making of the type proposed by Laibson and colleagues (e.g., Laibson, 1997; McClure, Laibson, Loewenstein, & Cohen, 2004). In that model, it is argued that the basal ganglia and medial prefrontal cortex form an emotional decision-making module that interacts (additively) with a second system organized around posterior parietal cortex and the dorsolateral prefrontal cortex, which form a rational decision-making module. Anatomical considerations that weigh against this hypothesis aside, we must ask whether or not there is compelling evidence that the division of brain areas into *competing* emotional and rational subgroups can be supported by the available data. In monkeys it has now been conclusively shown that activity in the posterior parietal cortex predicts preferences under all conditions that have been studied: for immediate rewards and for delayed rewards (Janssen & Shadlen, 2005; Louie & Glimcher, 2006), for large rewards and for small rewards (Platt & Glimcher, 1999; Dorris & Glimcher, 2004), for high-probability and low-probability rewards (Shadlen & Newsome, 1996; Platt & Glimcher, 1999). The data from animals seem to be unambiguous—lateral interparietal area (LIP) activity predicts choices for both rational and emotional decision making. To take another example, let us turn to the basal ganglia. This is an area a number of neuroeconomists have argued is associated with emotional decision making, but there is almost no evidence for this claim. Diseases of the basal ganglia are only very weakly associated with emotional dysfunction. The many dopaminergic forms of learning described here, although largely mediated by the basal ganglia, do not seem to capture any clear notion of emotionality. A similar case can be made for studies of the medial prefrontal cortex. As noted previously, there is evidence that this structure encodes monetary and primary rewards, preference, expected values, and gains and losses, and at least one study reports that it encodes long-delayed monetary gains. Together, these data paint a picture of structures globally involved in valuation driven by all mental states—not a structure driven exclusively by immediacy, fear, or emotionality.

In summary then, our available evidence seems to suggest that existing multiple-self models are largely unsupported by the bulk of our existing data. Of course, emotions influence decision making, and choosers show varying levels of self-control; those conclusions are beyond doubt. The question is, How do neural circuits related to emotions influence decision making? The amygdala, to take one example, may provide an answer. The amygdala projects strongly to the ventral striatum, and there is physiological and anatomical evidence that activity in the amygdala strongly influences activity in the ventral striatum. That evidence argues that the amygdala, and thus perhaps the emotions to which it is related, can influence valuation-related activity in this area. But it does not make a compelling case for a Freudian multiple-self model of neural decision making.

Choice

Unlike valuation, which has been extensively studied in both humans and other animals, choice has been the subject of study principally in awake behaving monkeys in neuroscience. That emphasis may reflect the fact that the temporal dynamics of choice make it difficult to study with fMRI. In any case, an understanding of choice requires an understanding of existing work in nonhuman primates.

Initial studies of choice in monkeys evolved almost simultaneously from studies of sensory-perceptual systems (e.g., Newsome, Britten, & Movshin, 1989) and movement-control studies (e.g., Glimcher & Sparks, 1992), as noted earlier. The most important of these studies examined how monkeys used noisy visual-sensory signals to identify one of two orienting eye movements, or saccades, as reinforced. They did so by leveraging an extensive preexisting literature on the structure of the visual and eye movement systems to search for the decision-making circuits that connected them in these tasks (Glimcher, 2003). Subsequent work has generalized many, but not all, of these findings to arm movement control systems and to studies of humans.

We have to begin, therefore, with a review of the basic structure of the saccadic control system (figure 75.2). The LIP in the posterior parietal cortex is one of the critical elements in this system, and it consists of a roughly topographic map both of objects in the visual world and the eye movements that would be required to align gaze with those objects (for a review see Glimcher, 2003). Thus a particular location on the map (or more precisely the neurons on the map at that location) activates when a visual stimulus appears 10 degrees to the right of fixation, *and* that region might become particularly active milliseconds before an eye movement that shifts gaze 10 degrees. This area, in turn, projects both to the frontal eye fields and the midbrain superior colliculus, two additional topographic maps that are broadly similar in function. The frontal eye fields project, as well, to the superior colliculus directly. A final note is that many of these areas are reciprocally connected (for a review of this anatomy, see Platt, Lau, & Glimcher, 2003), a fact which is probably important for understanding choice. Finally, the colliculus is connected to brain stem circuits that actually govern eye movements in real time. The connection between these brain stem systems and the colliculus are mediated by a class of collicular neurons called *burst neurons*. Burst neurons have the interesting biophysical property that they can fire action potentials in either of two states: a continuous low-frequency state in which many different firing rates are observed, and a burst state characterized by a fixed and extremely high firing rate.

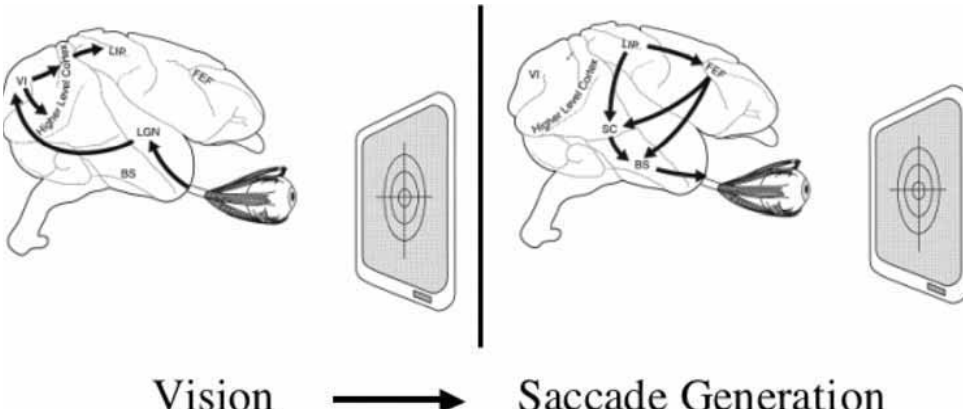


FIGURE 75.2 Saccadic control system: The visual-saccadic brain.

It is widely assumed that actual generation of a movement involves driving the collicular burst neurons above a specific firing-rate threshold, after which a burst occurs that is self-perpetuating and persists until the movement is complete. Inhibitory interconnections in the collicular map seem to preclude burstlike activity occurring at more than one location at a time, suggesting that the collicular architecture allows only a single movement to be executed at a time. Studies in area LIP, the frontal eye fields (FEF), and the superior colliculus (SC) all indicate that low-frequency firing in all three areas is related to the probability that a movement will be executed by the animal. To be more specific, if a particular movement is likely to yield a reward, then activity in all three maps at the locations associated with that movement is elevated. Of these three maps, the one that has been most studied with regard to decision is LIP. In LIP it has been shown that if the magnitude of a reward or the likelihood of a reward is systematically manipulated, then firing rates in these areas are a roughly linear function of those variables under many conditions (Dorris & Glimcher, 2004; Gold & Shadlen, 2007).

Together, these data suggest the following model for eye movement generation. At any moment in time neurons in LIP represent the instantaneous subjective value of each movement in the saccadic repertoire. Movements that have nonzero values are thus each represented by local activity on the map. One might even hypothesize that the representation of subjective value localized in the medial prefrontal cortex and the ventral striatum serves as the initial source of this signal.

In summary then, the available data suggest that all three of these areas, LIP, FEF, and SC, carry signals encoding subjective value and that movements occur when activity associated with one of the positively valued options drives its associated collicular neurons into their burst mode. A tremendous amount of work (reviewed in Glimcher, 2003; Gold & Shadlen, 2007) has examined this process of movement triggering under conditions in which animals are instructed to make movements as quickly as possible. Less is known about how movement selection is triggered in non-reaction-time settings. One important possibility is that an input to one or more of these areas alters the inhibitory interactions within the map, forcing convergence to a single action.

The basic model proposed for selecting eye movements is thus that signals encoding subjective value project to these areas, probably through LIP. These signals propagate recursively through these networks while reflecting value inputs that may be entering the maps at many locations. An external signal then permits, or forces, convergence of the network to a single choice that occurs when the collicular neurons are driven above their burst threshold.

Two questions, however, immediately arise: How does this system achieve choice among more abstract objects that do not have specific movements associated with them? Does this model generalize to humans and non-eye-movement conditions? A limited amount of data exist that do suggest that this general class of system operates under conditions in which choices are made between more abstract objects. Gold and Shadlen (2000; see also Sugrue, Corrado, & Newsome, 2004), for example, demonstrated that when animals must choose between red and green targets that constantly interchange locations, activity in the superior colliculus reflects the instantaneous mapping between color and value even if this changes from trial to trial. This finding clearly indicates that the saccadic choice circuit has access to instantaneous mapping information relating abstract properties to actions. It cannot tell us, however, how choice is accomplished (or if it can be accomplished) in the absence of any mapping to motor circuitry of any kind.

We do, however, have some interesting hints that these choice circuits are interconnected with important valuation areas in the frontal cortex and basal ganglia. Padoa-Schioppa and Assad (2006), for example, have demonstrated the existence of neurons in the orbitofrontal cortex that encode an animal's choice before the movement expressing that choice is executed. In a similar way, Lau and Glimcher (2008) have

observed choice neurons in the dorsal striatum. At the very least, this finding suggests that the choice circuit can send information about decisions frontally, but it may also indicate that these areas participate directly in the convergence process by which choice is accomplished.

The question of whether these circuits that have been so well studied in monkeys can be generalized to other classes of movements and other species is one about which we have much less information. We do know that adjacent to area LIP are areas specialized for arm, hand, and face movements. Standard theories (Andersen & Buneo, 2002) suggest that a group of areas lining the intraparietal sulcus serve as movement control interfaces for all of the body although there are problems still being resolved with those hypotheses (cf. Levy, Schluppeck, Heeger, & Glimcher, 2007). But it does seem clear that the general theories of eye movement control advanced for the monkey do have analogs in the skeletomuscular system. Further, injuries to any of these systems in either humans or monkeys lead to permanent deficits not in the musculature but in the ability to produce movements. Finally, a small number of fMRI studies have shown value-related signals in the posterior parietal cortex, although these signals are almost always of weaker magnitude than in more frontal areas. This result of course raises the possibility that the weaker fMRI signal reflects the temporal dynamics of choice. Because subjective value is only represented until a decision is made, in these areas the magnitude of the subjective value signal, integrated over an entire trial, may be much less than in areas located more frontally where subjective value is represented throughout a trial.

Summary

What emerges from a review of the available human and animal data on decision making is evidence of a two-stage model for choice. The first, or valuation, stage learns and represents the values of both actions and goods. Within this stage at least three learning mechanisms distributed in the basal ganglia and frontal cortex contribute to the construction of what we refer to as *subjective value*. These areas are hypothesized to learn subjective values, at a biophysical level, through the well-studied process of synaptic plasticity. These learning processes operate both during choice and during the passive receipt of rewards, effecting a disassociation between choice and valuation. A network, which includes the posterior parietal cortex and a number of movement-related areas subsequent to it in the motor control stream, appears to perform a winner-take-all operation on these values that accomplishes choice itself. Let me stress that the winner-take-all choice operation must be broadly distributed and involves structures that range from the superior colliculus to the orbitofrontal cortex.

Of particular interest are several features of the model that remain unspecified. While there are many candidate pathways by which information from the medial prefrontal cortex and the ventral striatum may influence activity in the posterior parietal cortex, which of these pathways is critical for choice has not yet been determined. We also have only limited information about the systems that "decide to choose." In some tasks, animals have to be trained to make a choice as soon as possible, and under these conditions one can observe the parietal and frontal networks converging rapidly toward choice. In other situations, however, the time courses of valuation and choice are separable. This possibility suggests the existence of a circuit that can essentially force the parietal networks toward convergence, the circuits that "decide to choose." Such circuits almost necessarily involve cortical networks of inhibitory connections, but the features of this process that decides when to choose remain completely absent from our standard model.

Over the course of the past decade a remarkable amount of progress has been made in identifying the basic features of the primate mechanism for choice. While many critical questions remain, progress in the last decade has marked this as an exciting and innovative area in cognitive neuroscience.

REFERENCES

Andersen, R. A., & Buneo, C. A. (2002). Intentional maps in the posterior parietal cortex. *Annu. Rev. Neurosci.*, *25*, 189–220.

Balleine, B. W., Daw, N. D., & O'Doherty, J. (2008). Multiple forms of value learning. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain*. New York: Elsevier.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*, 1–13.

Breiter, H. C., Aharon, D., Kahneman, D., Dale, A., & Shizgal, A. (2001). Functional imaging of neural responses to expectancy and experiences of monetary gains and losses. *Neuron*, *30*, 619–639.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychol. Rev.*, *58*, 313–323.

Dayan, P, & Balleine, B. W. (2002). Reward, motivation and reinforcement learning. *Neuron*, *36*, 285–298.

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.*, *84*, 3072–3077.

Dorris, M. C., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the subjective desireability of an action. *Neuron*, *44*, 365–378.

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *J. Neurosci*, *20*, 6159–6165.

Elliott, R., Newman, J. L., Longe, O. A., & Deakin, J. F. W. (2003). Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: A parametric functional magnetic resonance imaging study. *J. Neurosci.*, *23*, 303–307.

Glimcher, P. W. (2002). Decisions, decisions, decisions: Choosing a biological science of choice. *Neuron*, *36*, 323–332.

Glimcher, P. W. (2003). The neurobiology of visual saccadic decision making. *Annu. Rev. Neurosci.*, *26*, 133–179.

Glimcher, P. W., Camerer, C. F., Fehr, E., & Poldrack, R. A. (2008). *Neuroeconomics: Decision making and the brain.* New York: Elsevier.

Glimcher, P. W., Dorris, M. C., & Bayer, H. M. (2005). Physiological utility theory and the neuroeconomics of choice. *Games Econ. Behav.*, *52*, 213–256.

Glimcher, P. W., & Sparks, D. L. (1992). Movement selection in advance of action in the superior colliculus. *Nature*, *355*, 542–545.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature, 404*, 390–394.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*, 535–574.

Hikosaka, O. (2007). Basal ganglia mechanisms of reward-oriented eye movement. *Ann. NY Acad. Sci.*, *1104*, 229–249.

Hikosaka, O., Takikawa, Y., & Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive eye movements. *Physiol. Rev.*, *80*, 953–978.

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nat. Neurosci.*, *8*, 234–241.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.*, *10*, 1625–1633.

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. S., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: Characterization with rapid event-related fMRI. *NeuroImage*, *18*, 263–272.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *J. Neurosci.*, *25*, 4806–4812.

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage*, *12*, 20–27.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Q. J. Econ.*, *2*, 443–477.

Lau, B., & Glimcher, P. W. (2007). Action and outcome encoding in the primate caudate nucleus. *J. Neurosci.*, *27*, 14502–14514.

Lau, B., & Glimcher, P. W. (2008). Value representations in the primate caudate during a matching-law task. *Neuron.*, *58*, 451–463.

Levy, I., Schluppeck, D., Heeger, D. J., & Glimcher, P. W. (2007). Specificity of human cortical areas for reaches and saccades. *J. Neurosci.*, *27*, 4687–4696.

Ljungberg, T., Apicella, P., & Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *J. Neurophysiol.*, *67*, 145–163.

Louie, K., & Glimcher, P. W. (2006). Temporal discounting activity in monkey parietal neurons during intertemporal choice. In Society for Neuroscience Annual Meeting, Oct 14–18 2006. Atlanta, GA. Abstract nr 605.5.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503–507.

McClure, S., Li, J., Tomlin, D., Cypert, K., Montague, L., & Montague, P. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron, 44*, 379–387.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.*, *16*, 1936-1947.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future actions. *Nat. Neurosci.*, *9*, 1057–1063.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *341*, 52–54.

Newsome, W. T., Britten, K. H., Salzman, C. D., & Movshon, J. A. (1990). Neuronal mechanisms of motion perception. *Cold Spring Harb. Symp. Quant. Biol.*, *55*, 697–706.

Niv, Y., & Montague, P. R. (2008). Theoretical and empirical studies of learning. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the Brain.* New York: Elsevier.

O'Doherty, J. P., Buchanan, T. W., Seymour, B., & Dolan, R. J. (2006). Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron*, *49*(1), 157–166.

O'Doherty, J., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron*, *33*, 815–826.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*, 223–226.

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes in menu. *Nat. Neurosci.*, *11*, 95–102.

Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors in reward prediction. *Nat. Neurosci.*, *5*(2), 97–98.

Plassmann, H., O'Doherty, J., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.*, 27, 9984–9988.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*, 233–238.

Platt, M. L., Lau, B., & Glimcher, P. W. (2003). Situating the superior colliculus within the gaze control network. In W. C. Hall, & A. Moschovakis (Eds.), *The oculomotor system: New approaches for studying sensorimotor integration.* Boca Raton, FL: CRC Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black, & W. Prokasy (Eds.), *Classical conditioning II* (pp. 64-99). New York: Appleton-Century-Crofts.

Romo, R., & Schultz, W. (1990). Dopamine neurons of the monkey midbrain: Contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.*, *63*, 592–606.

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science*, *310*, 1337–1340.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.*, *13*, 900-913.

Schultz, W., & Romo, R. (1990). Dopamine neurons of the monkey midbrain: Contingencies of responses to stimuli eliciting immediate behavioral reactions. *J. Neurophysiol.*, *63*, 607–624.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: Seeing and deciding. *Proc. Natl. Acad. Sci. USA*, *93*, 628–633.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*, 1782–1787.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA: MIT Press.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science, 26*, 515–518.

76 Emotion and Decision Making

ELIZABETH A. PHELPS AND MAURICIO R. DELGADO

ABSTRACT The traditional characterization of emotion and decision making suggests that there are two opposing forces that can underlie choices: emotion, described as irrational and spontaneous, and reason, which is logical and deliberative. Although this dual-process approach is intuitively appealing, it fails to capture both the complexity of emotion and the diverse impact it has on decision processes. In this chapter, we highlight an integrative approach to understanding the relation between emotion and decision making. First, we review how two brain regions typically linked to decision making and emotion, the straitum and amygdala, both play roles in mediating value, learning, and action. Second, we review initial neuroeconomic studies that measure or manipulate specific emotion variables and examine their impact on decisions. Given that the purpose of emotion is to highlight what is relevant and important, it is not surprising that it should have a broad and critical role in decision making. However, research on the complex interaction of emotion and decision making is only beginning to emerge in neuroeconomics.

Historically, philosophical and scientific explorations of the relation between emotions and decisions have often described emotion as one system that may drive choices in contrast to reason (Dalgleish & Power, 1999). The notion that there are two opposing forces that can underlie choice behavior, with emotion being irrational and spontaneous, and reason logical and deliberative, has seeped into early research on neuroeconomics (e.g., Cohen, 2005). However, in spite of the intuitive appeal of this dual-systems approach, an initial examination of the neural circuitry of emotion and decision making reveals that choices are not so clearly "led by our heart or head." Rather, emotion is a nuanced concept that is composed of a number of overlapping and discrete processes (e.g., Scherer, 2005) that are not represented by a single neural system (LeDoux, 2000; Dalgleish, 2004). At the same time, several factors, representations, and neural systems can drive decisions (see chapter 74 by Rangel and chapter 75 by Glimcher, this volume), and an understanding of how these might interact with different components of emotion is only starting to emerge.

Aside from the folk psychological concept of emotion and reason as opposing forces in decisions, there are several other factors that have led to the slow development of an integrated understanding of the role of emotion in decision making and its underlying neural representation. First, decision science, adopting approaches from economics, has primarily investigated reward as the determinant of value in driving choice behavior. This emphasis has led to investigations of the neural systems of reward, as coded by dopamine. In cognitive neuroscience, this research has often emphasized the role of the striatum. In contrast, cognitive neuroscience research of emotion has been dominated by investigations of negative affect, specifically fear, perhaps because the characteristic physiological patterns of fear are relatively easy to investigate across species (Phelps, 2006). This research has focused on the role of the amygdala and classical fear conditioning, in which aversive values are learned, but are not tied to an action or choice (see LeDoux, this volume). Of course, choices are clearly driven by both appetitive and aversive values, and decision science will need to consider multiple reinforcer types to fully characterize decision making. Only by integrating research across the traditional disciplines of decision making and emotional learning can we begin to understand the complex relation between decisions and emotion. In the first part of this chapter, we will review the roles of the striatum and the amygdala in emotional learning, the coding of affective value, and the link to action.

A second factor that is only beginning to be addressed in studies of emotion and decision making is the complex conceptualization, organization, and measurement of emotion. Affective theorists and scientists recognize that emotion is not a unitary concept, but rather represents a class of processes, all of which are designed to signal the relevance or importance of events, situations, or information (e.g., Scherer, 2005). Although our ability to manipulate and measure emotion in the laboratory may not fully capture the range of factors outlined in theories of emotion, the suggestion that emotion has a single influence on decision making is quickly being revised (see Phelps, 2008, for a review). Along with this growing recognition of the complex psychological characterizations of emotion, affective neuroscientists have abandoned the idea that there is a single emotion "system" in the brain. This concept is the principle underlying the limbic system, which suggested that emotion is represented by a network of interacting neural structures (see Kotter & Meyer, 1992). However, in spite of years of investigations, there have yet to be definitive criteria for

ELIZABETH A. PHELPS Department of Psychology and Neural Science, New York University, New York, New York
MAURICIO R. DELGADO Department of Psychology, Rutgers University, Newark, New Jersey

inclusion in the limbic system (Kötter & Meyer, 1992), and the term has become more descriptively useful than scientifically informative (LeDoux, 2000). As a result, many affective neuroscientists have suggested that the limbic concept be abandoned, as it represents a simple but inaccurate view of emotion and the brain that might actually impede scientific progress as we try to uncover the complex neural representation of emotion (LeDoux, 2000; Phelps, 2008). In the second part of this chapter, we will briefly survey some initial findings from neuroeconomic research that have attempted to examine the complexity of emotion and isolate specific emotion variables and their influence on decisions.

The striatum and amygdala: Value, learning, and action

Striatum Overview Based on its connectivity and functional hetereogeneity, the striatum—the primary input structure of the basal ganglia—is centrally located to influence motivated or goal-directed behavior. The striatum receives convergent projections from the prefrontal cortex, along with afferents from various other key regions implicated in emotion such as the amygdala and the midbrain dopaminergic centers, projecting back to the cortex by way of the thalamus (Alexander & Crutcher, 1990; Graybiel, Aosaki, Flaherty, & Kimmura, 1994; Middleton & Strick, 1997; Graybiel, 2000). The striatum is anatomically divided into dorsal and ventral striatum in both rodents and primates (Lynd Balta & Haber, 1994; Gerfen & Wilson, 1996; Voorn, Vanderschuren, Groenewegen, Robbins, & Pennartz, 2004). In rodents, the dorsal striatum is further divided into a more dorsomedial component (roughly equivalent to the caudate nucleus in primates) and a more dorsolateral component (roughly equivalent to the putamen in primates) (Yin & Knowlton, 2006). The ventral striatum primarily refers to the nucleus accumbens (NAcc) in rodents, although the ventral portions of the putamen and caudate are also considered parts of the ventral striatum in primates based on connectivity. Functionally, dorsal and ventral distinctions in terms of affective processing are apparent irrespective of species (e.g., Robbins & Everitt, 1992; O'Doherty, 2004), but more recent accounts propose an integrated view of the striatum where information is processed in a gradient from more medial to lateral regions as learning progresses (Voorn et al.). Although traditionally thought of as a motor structure, over the years the basal ganglia, particularly the striatum, have gained notoriety for their involvement in more cognitive (Graybiel, 1995; Packard & Knowlton, 2002) and affective processes (Robbins & Everitt, 1992; O'Doherty, 2004) that contribute to goal-directed behaviors. Of central interest to this chapter is the role of the striatum in reward-related processing, which has gained support from a rich animal literature.

Research in the striatum across species demonstrates that it is not simply involved in encoding value as it is signaled by reward, but importantly in using reward signals to learn the stimuli and actions that predict reward, thus acquiring value and reinforcing adaptive actions. For example, electrophysiological recordings in the striatum in nonhuman primates highlight its involvement in the expectation of reward (Apicella, Scarnati, Ljungberg, & Schaltz, 1992; Kawagoe, Takikawa, & Hikosaka, 1998) and reward outcome processing (Hikosaka, Sakamoto, & Usui, 1989; Apicella, Ljungberg, Scarnati, & Schultz, 1991). This reward-response in the striatum is modulated according to properties of the stimulus such as the reward magnitude (Cromwell & Schultz, 2003) and individual preferences (Hassani, Cromwell, & Schultz, 2001). Similar to research in rodents, neurons in the primate dorsal striatum have been linked to the value of particular action that will lead to reward, rather than for the expected value of the outcome independent of action (Samejima, Ueda, Doya, & Kimura, 2005).

Consistent with electrophysiological data, deficits in approach or consummatory behaviors arise with lesions of the ventral and dorsal striatum, respectively (Robbins & Everitt, 1992). While the rodent ventral striatum has been implicated in simple associative reward-related learning, as exemplified by Pavlovian or classical conditioning (Ito, Dalley, Howes, Robbins, & Everitt, 2000), research suggests that the rodent dorsal striatum is more involved in learning the value of action outcomes, as exemplified by instrumental conditioning (Ito, Dalley, Robbins, & Everitt, 2002), with a potential gradient suggesting dorsomedial regions mediating initial acquisition and dorsolateral regions involved with the more habitual and automatic motor aspects of behavior with extensive learning (Yin, Knowlton, & Balleine, 2005; Yin, Ostlund, Knowlton, & Balleine, 2005).

The contributions of the striatum to reward processing have more recently been extended to the human brain, given the rapid advances in functional magnetic resonance imaging (fMRI). Many of the initial studies quickly highlighted the role of the striatum in expectation or anticipation of primary rewards such as juice (Berns, McClure, Pagnoni, & Montague, 2001; O'Doherty, Deichmann, Critchley, & Dolan, 2002), secondary more abstract rewards such as money (Knutson, Adams, Fong, & Hommer, 2001; Kirsch et al., 2003), and even processes associated with maladaptive reward expectations such as drug craving (Breiter et al., 1997). Responses to rewarding outcomes were also reported in the striatum (Delgado, Nystrom, Fissell, Noll, & Fiez, 2000), with variations in magnitude and probability of rewards (Delgado, Locke, Stenger, & Fiez, 2003; Delgado, Miller, Inati, & Phelps, 2005; Galvan et al., 2005; Nieuwenhuis et al., 2005) irrespective of whether the reward was extrinsic (e.g., monetary) or intrinsic (e.g., positive or negative feedback on cognitive tasks; Elliott, Sahakian,

Michael, Paykel, & Dolan, 1998; Poldrack et al., 2001; Seger & Cincotta, 2005). Thus early neuroimaging research was able to support and extend findings from nonhuman animals and implicate the striatum as an important structure in processing information about incentives that influence motivated behavior.

Given that one of the primary projections to the striatum is from dopaminergic neurons in the midbrain (Haber, Fudge, & McFarland, 2000) and that blood oxygenation level dependent (BOLD) responses in the striatum may reflect synaptic input activity (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001), the human striatum hemodynamic responses were considered to reflect more than just anticipatory and consummatory signals in the brain. Rather, they were suggested to reflect a learning and valuation signal mediated by human corticostriatal circuits that underlies goal-directed or motivated behaviors (Montague & Berns, 2002; Balleine, Delgado, & Hikosaka, 2007). It was soon observed that prediction errors—the difference between expected and experienced affective outcomes thought to be coded by dopamine neurons in nonhuman primates (Schultz, Dayan, & Montague, 1997; Schultz & Dickinson, 2000; Bayer & Glimcher, 2005)—were coded in the human striatum during simple conditioning experiments with primary rewards, such as juice (Pagnoni, Zink, Montague, & Berns, 2002; McClure, Berns, & Montague, 2003; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003). The finding that prediction errors correlate with human striatal activity is widely reported with different types of paradigms and reinforcers (e.g., Schönberg, Daw, Joel, & O'Doherty, 2007; Tobler, O'Doherty, Dolan, & Schultz, 2007; Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008). Additionally, in support of the animal literature, research suggested that the human striatum response to outcomes reflected reinforcement learning signals, rather than a reward per se (Tricomi, Delgado, & Fiez, 2004). Interestingly, while the human ventral striatum is involved in reward predictions irrespective of learning paradigm (e.g., classical or instrumental conditioning; O'Doherty et al., 2004), the dorsal striatum has been linked to more action-contingent learning (O'Doherty et al., 2004; Tricomi et al.; Zink, Pagnoni, Martin-Skurski, Chappelow, & Berns, 2004) with diminishing responses as learning progresses and stimuli become more predictable (Haruno et al., 2004; Delgado, Miller, et al., 2005; Seger & Cincotta, 2005).

Currently, neuroscience research on the striatum's role in reward processing has merged with different disciplines, such as economics (Glimcher & Rustichini, 2004) and social psychology (Ochsner, 2004; Frith & Singer, 2008), to examine the different valuation calculations that contribute to motivated behavior and how such calculations are influenced by everyday social factors. Neuroeconomic investigations, for instance, have suggested that complex calculations of expected or subjective value are integrated in the striatum during dynamic tasks that require constant representations and updating of reward contingencies, representing magnitude and probability (Knutson, Taylor, Kaufman, Peterson, & Glover, 2005), along with risk possibilities (Hsu, Blatt, Adolphs, Tranel, & Camerer, 2005; Schultz et al., 2008) and even information regarding time (McClure, Laibson, Loewenstein, & Cohen, 2004; Kable & Glimcher, 2007) and effort (Botvinick, Huffstetler, & McGuire, 2009). Although investigations with respect to losses and negative values in the striatum are in their infancy, they also have to be factored into these calculations (Becerra, Breiter, Wise, Gonzalez, & Borsook, 2001; Tom, Fox, Trepel, & Poldrack, 2007; Delgado, Li, Schiller, & Phelps, 2008; Delgado, Schotter, Ozbay, & Phelps, 2008).

Social factors also have to be taken into account when considering the role of the striatum in reward-related processing. Learning to trust someone, for instance, is a trial-and-error procedure in which feedback attained during social interaction is crucial for developing the feeling of trust. When participants develop reputations through experience in a trust game (King-Casas et al., 2005), neural signals in the striatum shift according to learning; that is, they respond to the earliest predictor of a potential reward—a pattern analogous to responses observed in dopamine models of reinforcement learning. Similarly, social signals in the trust game (such as faces of previous cooperators) can serve as approach signals for future social interactions and engage striatal circuits (Singer, Kiebel, Winston, Dolan, & Frith, 2004) that will be motivated to act in self-interest (e.g., retaliation) when trust is breached (de Quervain et al., 2004). Thus social information can modulate striatal function involved in reward processing, particularly when social expectations are violated (Delgado, Frank, et al., 2005). Overall, research on the striatum has highlighted its role in processing a wide range of rewards and the use reward signal to learn the value of stimuli and reinforce adaptive actions.

Amygdala Overview The amygdala is a small, almond-shaped structure in the medial temporal lobe that sits adjacent and anterior to the hippocampus and has broad connections with sensory, prefrontal, and subcortical regions (M. Young, 2002). In spite of its small size, the amygdala is a complex structure with several subnuclei, each with unique roles. The primary technique used to explore the function of these subnuclei in animal models is classical fear conditioning, in which a neutral stimulus, such as a tone, called the conditioned stimulus (CS), is paired with aversive stimulus, such as a shock, called the unconditioned stimulus (US). Through this pairing, the CS comes to elicit a range of conditioned fear responses (CR) when presented alone. Within the amygdala, the lateral nucleus is the site of sensory input from the CS and US, and is also proposed to be the

site of plasticity underlying the CS-US association (see LeDoux this volume). The LA projects to the central nucleus (CE), which in turn outputs to a range of brain regions mediating the physiological and neurohormonal expression of conditioned fear. The LA also projects to the basal nucleus (B), which projects the CE and plays a role in the contextual mediation of the expression of fear (Fanselow, 2000).

The amygdala's role in fear conditioning highlights its importance in the encoding and storage of the emotional value of stimuli and the physiological expression of fear, but this research alone does not link the amygdala to any action or choice. Although it has been proposed that the physiological expression of emotion may underlie decisions, perhaps through interactions with orbitofrontal cortex (Damasio, 1994), recent research in rodents suggests that the amygdala may mediate choice more directly. A study exploring the amygdala's role in mediating choice examined the escape from fear (EFF) paradigm. In this paradigm, the rodent first undergoes fear conditioning. In a second stage, the rodent is given the option to take an action to terminate the CS, thus reducing exposure to the fear-eliciting event. The termination of the CS becomes a conditioned reinforcer for the instrumental action. Amorapanth, LeDoux, and Nader (2000) found that diminishing fear through an action or choice relied on a circuitry within the amygdala that could be dissociated from its physiological expression. By placing lesions in the different amygdala subnuclei, it was found that although damage to the LA impairs both the physiological expression of fear and learning an action to diminish exposure to the CS, damage to the CE only disrupted the physiological expression. Rodents with lesions confined to the CE were able to learn an action to terminate the CS, even though they failed to show the typical expression of conditioned fear. In contrast, damage to B resulted in the opposite pattern of results, that is, failure to learn an action to terminate the CS but normal expression of the CR. It is suggested that B may not be the site of storage for the representation of action or choice in the EFF paradigm; rather, B projects to the striatum, which, as outlined earlier, has a broader role in motor control and reinforcement of action. In short, the pathway for fear-motivated actions is hypothesized to involve the LA, which projects to the B, which, in turn, projects to the striatum to convey the reinforcing nature of the instrumental action (see LeDoux & Gorman, 2001).

Because of technical limitations, research on the human amygdala has not yet examined the roles of specific subnuclei, but in general the function of the amygdala appears to be similar across species. Imaging studies demonstrate increased amygdala activation to a CS that is correlated with expression of the CR (LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998; Buchel, Morris, Dolan, & Friston, 1998). Damage to the human amygdala impairs the physiological expression of conditioned fear (Bechara et al., 1995; LaBar, LeDoux, Spencer, & Phelps, 1995). These results suggest that the mechanism of fear conditioning is preserved across species. However, humans have more efficient means of learning which stimuli in the environment predict potential aversive consequences than fear conditioning. Learning fears through social means does not require the aversive experience of a US. For instance, simply being told or instructed that an event or situation may lead to an aversive consequence is enough to elicit a fear response in that situation (Funayama, Grillon, Davis, & Phelps, 2001: Phelps et al., 2001). Similarly, observing another person being hurt when exposed to a stimulus is enough to elicit a fear when that stimulus is presented (Olsson, Nearing, & Phelps, 2007). Social learning of fear is more typical in humans, and learning through language is unique to humans. However, research indicates that even uniquely human, socially acquired fears also depend on the phyologentically old amygdala for expression (see Olsson & Phelps, 2007, for a review). Finally, there is some preliminary evidence suggesting that amygdala-striatal interactions may also play a role in mediating actions that diminish exposure to fearful events in humans (Delgado, Jou, LeDoux, & Phelps, 2007), consistent with animal models.

Aside from emotional learning, the amygdala has other roles that are less explicitly linked to decisions or actions, but could have an impact on the assessment of value. One such role is to modulate cognition in the presence of emotional events. Because of its extensive connectivity throughout the brain, the amygdala is ideally situated to detect emotional events in the environment quickly and enhance further processing to assure that these events receive priority. For example, through its interaction with the hippocampus, the amygdala enhances the storage or consolidation of memories with arousal to assure that emotional events persist in memory (McGaugh, 2000). In addition, the amygdala's reciprocal connection with sensory cortices results in the facilitation of attention for emotional stimuli (Anderson & Phelps, 2001) and enhanced sensory cortical responses (Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004). By representing some aspects of affective value, the amygdala, through its wide connectivity, can insure that cognition is tuned to events with affective significance.

Finally, the amygdala's role extends to representing some aspects of social value. This is most evident in the detection of fear from facial expressions (Adolphs et al., 1999). Patients with lesions to the amygdala fail to rate faces with fear expressions as fearful, a deficit that seems primarily driven by their tendency to ignore the eyes, relative to normal control subjects (Adolphs et al., 2005). The amygdala is also involved in judging faces as more or less trustworthy (Adolphs, Tranel, & Damasio, 1998) and responds to some of the evaluative qualities of social group membership (Phelps et al., 2001). This social role for the amygdala includes some

aspects of detecting social intent. When viewing a movie in which geometric shapes move around in a way that suggests a social interaction, control subjects report a social interaction, whereas patients with lesions to the amygdala only report the movement of shapes (Heberlein & Adolphs, 2004). These broader roles for the amygdala in modulating cognitive processes and representing aspects of social value could affect decisions in a number of ways that have yet to be explored.

THE STRIATUM, THE AMYGDALA, AND VALUE In the preceding discussion, we highlighted a role for both the striatum and the amygdala in representing value, but how we define value varies. With the striatum, we discuss value as represented by reward signals in response to stimuli that would normally cause one to approach a situation, such as juice if one is thirsty, money, or social cooperation. With the amygdala, we refer to value representing an emotional significance that would cause one to avoid a situation, such as fear or untrustworthiness. This description suggests a clear dichotomy as to unique roles of the striatum and amygdala in the representation value. However, the data are not nearly as clear as the simple description suggests.

In fact, there is abundant evidence that the role of the striatum is not strictly limited to representing reward and appetitive learning and the amygdala is not solely involved in aversive conditioning. Rather, both structures play a role in mediating appetitive and aversive learning. For example, in the NAcc, dopamine responses are elevated in response to aversive outcomes, such as electric shocks, tail pinch, and social stress (Robinson, Becker, Young, Akil, & Castaneda, 1987; Kalivas & Duffy, 1995), and also in response to a CS or a context predicting potential aversive consequences (Murphy, Pezze, Feldon, & Heidbreder, 2000; A. Young, 2004). Dopamine in the NAcc is also important for aversive instrumental conditioning, such as active or passive avoidance and escape tasks (Schwarting & Carey, 1985; Wadenberg, Ericson, Magnusson, & Ahlenius, 1990; McCullough, Sokolowski, & Salamone, 1993). The dorsal striatum is also linked to aversive learning, with lesions to this region resulting in deficits in conditioned fear and active avoidance (Winocur & Mills, 1969; White & Salinas, 2003).

Consistent with research in other species, human fMRI studies also implicate the striatum in aversive classical and instrumental conditioning. Even though the striatum is not often the focus of neuroimaging studies on fear conditioning, most of these studies report activation of the striatum, in addition to the amygdala (Buchel et al., 1998; LaBar et al., 1998; Buchel, Dolan, Armony, & Friston, 1999; Phelps, Delgado, Nearing, & LeDoux, 2004). Striatal activation has also been observed in relation to pain (Ploghaus et al., 2000; Seymour et al., 2005) and monetary loss (Delgado, 2007; Seymour, Daw, Dayan, Singer, & Dolan, 2007; Tom et al., 2007). In a recent review, it was found that across aversive learning paradigms, the striatum consistently represents prediction error signals, whereas the amygdala does not, suggesting a general role for the striatum in mediating temporal difference learning in both appetitive and aversive tasks (Delgado, Li, et al., 2008).

As outlined earlier, the amygdala is critical for classical fear conditioning, but it has also been suggested to play a role in classical conditioning paradigms involving rewards. It has been suggested that the basolateral nucleus of the amygdala (BLA) is important for maintaining and updating the representation of the affective value of an appetitive CS (Parkinson et al., 2001), which may rely on its interactions with the corticostriatum and dopaminergic circuitry (Rosenkranz & Grace, 2002; Rosenkranz, Moore, & Grace, 2003). Given this possibility the BLA appears to be involved in variations of standard appetitive conditioning paradigms, including when using a secondary reinforcer, when the value of the US has changed after conditioning, or in tasks that explore the interactions between Pavlovian and instrumental processes (Gallagher, Graham, & Holland, 1990). Similar effects have been observed when the BLA is disconnected from the NAcc, suggesting that amygdala-striatal interactions are important for processing learned motivational value (Setlow, Holland, & Gallagher, 2002). In addition, the CE may be critical for some expressions of appetitive conditioning, including enhanced attention (orienting) to the CS (Gallagher et al.), and controlling the general motivational influence of rewarding events (Corbit & Balleine, 2005). There is some evidence in humans supporting a role for the amygdala in appetitive conditioning. Patients with amygdala lesions are impaired in conditioned preference tasks (Johnsrude, Owen, White, Zhao, & Bohbot, 2000), and neuroimaging studies report amygdala activation during appetitive conditioning with food as a US (Gottfried, O'Doherty, & Dolan, 2003).

Although the complex interactions between the amygdala and striatum in affective learning and decisions needs to be further clarified, these studies indicate that the striatum and amygdala interact to represent learned value for both appetitive and aversive reinforcers, suggesting that there is not a clear division between the neural structures typically linked to value in decision making and those representing emotional value. While neuroeconomic researchers generally distinguish between emotion and value when investigating decision processes, affective scientists have a much broader view of value as a fundamental component of emotion. Future research integrating emotion and decision making will have to grapple with are the precise definition of value across disciplines, as well as the complex interactions of the striatum and amygdala in mediating value, learning, and actions.

Initial investigations of emotion's influence on economic decision making

In this section, we briefly review some published neuroeconomic research that has explicitly attempted to incorporate emotion variables. As mentioned earlier, emotion encompasses several independent and overlapping processes and functions that share a common goal in signaling relevance or importance to the organism (Frijda, 1986; Scherer, 2005). In spite of a growing interest in the role of emotion in decision making (e.g., Damasio, 1994; Cohen, 2005), relatively few studies have actually attempted to manipulate or measure specific emotion variables. We will highlight a few that have done so and outline what aspect of emotion each study incorporates. We have not included in this overview studies that use BOLD responses or lesions in specific brain regions as the primary assessment or manipulation of emotion. Although it is clear that certain brain regions have emotion functions, almost all these regions have roles in cognitive or sensory processes as well, so it hard to infer emotion directly from BOLD or a lesion in the absence of any additional assessment of emotion (Phelps, 2008).

Assessing Emotion and Linking It to Choice There are several means to assess emotion, including subjective report, physiological measurement, and actions (see Phelps, 2008, for a review). With the exception of action, which is a general response that is not unique to emotion, very few neuroeconomic studies have attempted to assess emotional responses and relate them to decisions. One exception is the classic study by Bechara, Damasio, Tranel, and Damasio (1997) in which they found that physiological arousal was correlated with decisions in an economic gambling task.

Using the Iowa Gambling Task (IGT) participants were required to pick a card from one of four decks over a series of trials. Each card represented a monetary reward or punishment, and the goal was to learn through trial and error which deck would yield a higher profit over time. The decks varied in their payoff schedules, with two of the decks yielding greater rewards, but also greater punishments resulting in less overall profit than the other two decks. Over trials, normal, healthy participants learned to increase their profits by selecting from the more advantageous decks more often than the less advantageous decks. Physiological arousal was assessed as participants contemplated selecting from the decks. Prior to selecting from the less advantageous decks, normal participants showed increased arousal. It was suggested that this arousal response cued the participant to move away from those decks and toward a more advantageous choice. In short, it was argued that physiological arousal mediated decision making in this task. Bechara and colleagues (1997) also conducted this study in patients with damage to the orbitofrontal cortex. These patients failed to learn to select more from the advantageous decks over time and also failed to generate anticipatory arousal responses. These results inspired Antonio Damasio (1994) to propose the *somatic marker hypothesis*, which suggests that bodily states and emotional responses play a fundamental role in driving choice behavior.

Although several additional studies over the years have failed to find strong support for the somatic marker hypothesis (see Dunn, Dalgleish, & Lawrence, 2006, for a review), the Bechara and colleagues (1997) study represents one of the first attempts to assess an emotional response and relate it to decisions. In this way, it highlighted the importance of the considering emotion in decision-making research and inspired future investigations on this topic.

Manipulating Emotion Directly and Observing the Impact on Choice There are several means to manipulate emotion in the laboratory (Phelps, 2008), but only a few of these techniques have been used in economic or neuroeconomic research. One of the first neuroeconomic studies to manipulate emotion and examine its impact on decisions utilized a pharmacological manipulation. Oxytocin is a drug that increases social affiliation and the social emotion of trust. Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr (2005) administered either oxytocin or placebo to participants and then asked them to play the trust game. In this game, the investor has the opportunity to share a variable sum of money with a trustee. By sharing this money the overall amount is magnified. The trustee has the option to share back some of the profits or keep all the money for him- or herself. Thus sharing by the investor is risky and requires trust. The oxytocin group showed much higher rates of trust (and risk) by sharing more money, relative to the placebo. However, oxytocin did not increase risk in a nonsocial game, only in the social interaction. Although this initial study did not investigate the underlying neural systems mediating increased trust, a follow-up study found that responses in both the amygdala and dorsal striatum were linked to differential choices in the oxytocin group (Baumgartner, Heinrichs, Vonlantehn, Fischbacher, & Fehr, 2008).

In affective science, a more traditional means to manipulate emotion is to present stimuli that elicit an emotional response. Although there are no neuroeconomic studies we know of that have used this approach, there are a few behavioral economic studies. For instance, in a study aimed at investigating the influence of acute stress on financial decision making, stress was induced by asking participants to immerse their hand for a period of time in either cold (stress) or room-temperature (no stress) water (Porcelli & Delgado, 2009). It was hypothesized that the physiological and emotional consequences of stress would interfere with deliberative decision-making processes, leading one to depend on more intuitive or automatic processes. Participants were

asked to choose between two gambles that varied in terms of risk and value. Importantly, the choice between a risky and a conservative gamble was presented in a gain domain (the outcomes of both gambles are positive) and a loss domain (the outcomes of both gambles are negative). Participants' choices conformed to a pattern of behavior known as the reflection effect—that is, greater preference for risky options when decisions involve losses rather than gains (Kahneman & Tversky, 1979). This bias was exacerbated when participants were under stress. That is, participants were more conservative when choosing between gambles with positive outcomes (gain), while being riskier when choosing between gambles with negative outcomes (loss). This study suggests that participants perhaps fall back on automatized reactions to risk when in a stressful environment.

Another means to manipulate affect is to alter participants' moods. Mood can elicit action tendencies to approach or avoid (Scherer, 2005) that may impact economic choices. Lerner, Small, and Loewenstein (2004) induced sad and disgust moods and examined their impact on a classic economic problem—the endowment effect. When they came into the lab, half the participants were given a highlighter set. All the participants then watched one of three movies, chosen to elicit a sad, disgust, or neutral mood. Along with viewing the movie, participants were asked to write about how they would feel if they experienced the events depicted in the movie. After this procedure, participants were asked how much they would sell the highlighter set for (if endowed) or choose to pay for it (if not endowed). Those in the neutral mood condition showed the classic endowment effect—they would demand more to sell the product than they would pay for it. In contrast, the sad group showed the opposite. They were willing to sell the highlighter set for a lower price than they would pay for it. The disgust group showed no endowment effect at all. These results demonstrate not only that mood can have an effect on economic choices, but also that the effect is specific to different mood states.

Manipulating Emotion Through Appraisal and Observing Its Impact on Choice A fundamental component of emotion is appraisal (see Scherer, 2005, for a review). Appraisal is the process that occurs when an event is first encountered and its emotional meaning is determined. Although the evaluation or appraisal of the emotional significance of an event can occur quickly and automatically, appraisals and emotional reactions can also unfold as information about the circumstances is acquired. A classic example of the importance of appraisals in generating emotion is a study by Schacter and Singer (1962). They evoked physiological changes consistent with emotion by administering epinephrine. However, they only informed a subset of participants that the bodily changes are related to the drug. All the participants were then placed in a social circumstance that might evoke happiness or anger. Only those participants who were unaware of the cause of their physiological state reported the subjective experience of euphoria or anger and behaved in a manner consistent with these emotions. This classic finding highlights the importance of appraisal in generating an emotional response.

In affective science, several types of manipulations or instruction have been shown to influence appraisal and subsequent emotional reactions (e.g., Ochsner & Gross, 2008). In economics, one such manipulation might be framing. By emphasizing loss or gain in how an identical economic choice is presented, participants will make a different decision in order to avoid loss (Kahneman & Tversky, 1979). It is suggested that this framing manipulation, by changing appraisal, alters the emotional significance of the outcome, thus biasing the decision (Gilovich, Griffin, & Kahneman, 2002). A study examining the neural basis of the framing effect found that BOLD responses in the amygdala were significantly altered by the framing manipulation (DeMartino, Kumaran, Seymour, & Dolan, 2006). This finding is consistent with a large body of research on emotion regulation suggesting that altering appraisal can influence amygdala activity and a range of emotional responses (see Ochsner & Gross, 2008, for a review). In addition, activity in the orbital and medial prefrontal cortices predicted the susceptibility to framing effects across individuals (DeMartino et al.).

A second study examined the impact of framing on economic auctions (Delgado, Schotter, et al., 2008). This study specifically investigated why people tend to pay too much, or overbid, in experimental auctions. Using fMRI, it was found when playing a two-person auction game versus a lottery game against a computer, the striatum showed an exaggerated response when the auction was lost, even though there was no monetary loss. Correlating the striatal response to winning or losing the auction with the bid chosen showed that only the loss response was related to the bid chosen and the tendency to overbid. This finding led to the hypothesis that perhaps it is the contemplation or fear of loss, in this case social loss, that is driving the tendency to overbid. To explore this hypothesis, a behavioral framing study was conducted in which some participants were given $15 and were told they would lose it if they lost the auction, whereas other participants were told they would receive a bonus of $15 if they won the auction. In these equivalent auctions, the loss frame increased overbidding and profit for a hypothetical auctioneer.

These framing studies (DeMartino et al., 2006; Delgado, Schotter, et al., 2008) produced results consistent with the suggestion that altering the appraisal of the economic choice can influence emotion and alter the decision. However, neither of these studies can confirm, with a direct assessment, that emotion was altered. A recent behavioral study adapted an emotion-regulation technique to examine the

impact of altering appraisal on both economic decisions and emotional reactions (Sokol-Hessner et al., 2009). In this study, participants were presented with risky gambles and had to choose between the gamble and a guaranteed outcome. On half of the trials, participants were instructed to view each choice as one in a portfolio of choices. For the other half, they were instructed to focus on each choice as if it were the only choice. Using the portfolio frame resulted in an overall decrease in loss aversion. The more loss-averse a participant was, the higher the physiological arousal response to monetary losses relative to gains at outcome. Furthermore, those participants who were more successful at decreasing loss aversion by adopting the portfolio frame also showed a greater decrease in their arousal response to losses. These findings indicate that a classic economic phenomenon, loss aversion, may be related to an emotional response, arousal, and that changing the framing or appraisal of a choice may impact both its emotional impact and loss aversion.

Conclusion

The purpose of emotion is to signal to the organism the events, situations, or information in the environment that may be relevant and potentially important for survival (Scherer, 2005). Therefore, it is not surprising that emotion should have a broad and critical role in decision making. However, research on the interaction of emotion and decision making is only beginning to emerge in neuroeconomics. Although it has long been acknowledged that emotion affects decisions (Kahneman, 2003), the traditional view that emotion is a unitary construct that has a single impact on decision processes fails to embrace the complexity of emotion and the broad and diverse impact it can have on decision processes across domains. As neuroeconomics begins to adopt the tools and theories of affective science and affective neuroscience, a more integrated view of decision making will emerge.

REFERENCES

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*, 68–72.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*, 470–474.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., Anderson, A., Lee, G. P., & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*, 1111–1117.

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends Neurosci.*, *13*, 266–271.

Amorapanth, P., LeDoux, J. E., & Nader, K. (2000). Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nat. Neurosci.*, *3*, 74–79.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*, 305–309.

Apicella, P., Ljungberg, T., Scarnati, E., & Schultz, W. (1991). Responses to reward in monkey dorsal and ventral striatum. *Exp. Brain Res.*, *85*, 491–500.

Apicella, P., Scarnati, E., Ljungberg, T., & Schultz, W. (1992). Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *J. Neurophysiol.*, *68*, 945–960.

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J. Neurosci.*, *27*, 8161–8165.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation. *Neuron*, *58*, 639–650.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*, 129–141.

Becerra, L., Breiter, H. C., Wise, R., Gonzalez, R. G., & Borsook, D. (2001). Reward circuitry activation by noxious thermal stimuli. *Neuron*, *32*, 927–946.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*, 1293–1295.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, *269*, 1115–1118.

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *J. Neurosci.*, *21*, 2793–2798.

Botvinick, M. M., Huffstetler, S., & McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cogn. Affective Behav. Neurosci.*, *9*(1), 16–27.

Breiter, H. C., Gollub, R. L., Weisskoff, R. M., Kennedy, D. N., Makris, N., Berke, J. D., Goodman, J. M., Kantor, H. L., Gastfriend, D. R., Riorden, J. P., Mathew, R. T., Rosen, B. R., & Hyman, S. E. (1997). Acute effects of cocaine on human brain activity and emotion. *Neuron*, *19*, 591–611.

Buchel, C., Dolan, R. J., Armony, J. L., & Friston, K. J. (1999). Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. *J. Neurosci.*, *19*, 10869–10876.

Buchel, C., Morris, J., Dolan, R. J., Friston, K. J. (1998). Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron*, *20*, 947–957.

Cohen, J. D. (2005). The vulcanization of the human brain: A neural perspective on interactions between cognition and emotion. *J. Econ. Perspect.*, *19*, 13–24.

Corbit, L. H., & Balleine, B. W. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of Pavlovian-instrumental transfer. *J. Neurosci.*, *25*, 962–970.

Cromwell, H. C., & Schultz, W. (2003). Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J. Neurophysiol.*, *89*, 2823–2838.

Dalgleish, T. (2004). The emotional brain. *Nat. Rev. Neurosci.*, *5*, 582–589.

Dalgleish, T., & Power, M. J. (1999). *Handbook of cognition and emotion*. New York: Wiley.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason and the human brain*. New York: G. P. Putnam's Sons.

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Ann. NY Acad. Sci.*, *1104*, 70–88.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005) Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.*, *8*, 1611–1618.

Delgado, M. R., Jou, R. L., LeDoux J., & Phelps, E. A. (2007). Avoiding negative outcomes: Tracking the mechanisms of avoidance learning in humans during fear conditioning. Program No. 97.1, *Soc. Neurosci.*

Delgado, M. R., Li, J., Schiller, D., & Phelps, E. A. (2008). Review. The role of the striatum in aversive learning and aversive prediction errors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *363*, 3787–3800.

Delgado, M. R., Locke, H. M., Stenger, V. A., & Fiez, J. A. (2003). Dorsal striatum responses to reward and punishment: Effects of valence and magnitude manipulations. *Cogn. Affect. Behav. Neurosci.*, *3*, 27–38.

Delgado, M. R., Miller, M. M., Inati, S., & Phelps, E. A. (2005). An fMRI study of reward-related probability learning. *NeuroImage*, *24*, 862–873.

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.*, *84*, 3072–3077.

Delgado, M. R., Schotter, A., Ozbay, E. Y., & Phelps, E. A. (2008). Understanding overbidding: Using the neural circuitry of reward to design economic auctions. *Science*, *321*, 1849–1852.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases and rational decision-making in the brain. *Science*, *313*, 684–687.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–1258.

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neurosci. Biobehav. Rev.*, *30*, 239–271.

Elliott, R., Sahakian, B. J., Michael, A., Paykel, E. S., & Dolan, R. J. (1998). Abnormal neural response to feedback on planning and guessing tasks in patients with unipolar depression. *Psychol. Med.*, *28*, 559–571.

Fanselow, M. S. (2000). Contextual fear, gestalt memories, and the hippocampus. *Behav. Brain Res.*, *110*, 73–81.

Frijda, N. H. (1986). *The emotions*. Cambridge, UK: Cambridge University Press.

Frith, C. D., & Singer, T. (2008). Review. The role of social cognition in decision making. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *363*, 3875–3886.

Funayama, E. S., Grillon, C., Davis, M., & Phelps, E. A. (2001). A double dissociation in the affective modulation of startle in humans: Effects of unilateral temporal lobectomy. *J. Cogn. Neurosci.*, *13*, 721–729.

Gallagher, M., Graham, P. W., & Holland, P. C. (1990). The amygdala central nucleus and appetitive Pavlovian conditioning: Lesions impair one class of conditioned behavior. *J. Neurosci.*, *10*, 1906–1911.

Galvan, A., Hare, T. A., Davidson, M., Spicer, J., Glover, G., & Casey, B. J. (2005). The role of ventral frontostriatal circuitry in reward-based learning in humans. *J. Neurosci.*, *25*, 8650–8656.

Gerfen, C. R., & Wilson, C. J. (1996). The basal ganglia. In L. W. Swanson, A. Bjorklund, & T. Hokfelt (Eds.), *Handbook of chemical neuroanatomy* (pp. 371–468). Amsterdam: Elsevier Science.

Gilovich, T., Griffin, D. W., & Kahneman, D. (Eds.) (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

Glimcher, P. W., & Rustichini, A. (2004). Neuroeconomics: The consilience of brain and decision. *Science*, *306*, 447–452.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, *301*, 1104–1107.

Graybiel, A. M. (1995). Building action repertoires: Memory and learning functions of the basal ganglia. *Curr. Opin. Neurobiol.*, *5*, 733–741.

Graybiel, A. M. (2000). The basal ganglia. *Curr. Biol.*, *10*, R509–511.

Graybiel, A. M., Aosaki, T., Flaherty, A. W., & Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science*, *265*, 1826–1831.

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J. Neurosci.*, *20*, 2369–2382.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.*, *28*, 5623–5630.

Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., Imamizu, H., & Kawato, M. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: A functional magnetic resonance imaging study of a stochastic decision task. *J. Neurosci.*, *24*, 1660–1665.

Hassani, O. K., Cromwell, H. C., & Schultz, W. (2001). Influence of expectation of different rewards on behavior-related neuronal activity in the striatum. *J. Neurophysiol.*, *85*, 2477–2489.

Heberlein, A. S., & Adolphs, R. (2004). Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proc. Natl. Acad. Sci. USA*, *101*, 7487–7491.

Hikosaka, O., Sakamoto, M., & Usui, S. (1989). Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J. Neurophysiol.*, *61*, 814–832.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science*, *310*, 1680–1683.

Ito, R., Dalley, J. W., Howes, S. R., Robbins, T. W., & Everitt, B. J. (2000). Dissociation in conditioned dopamine release in the nucleus accumbens core and shell in response to cocaine cues and during cocaine-seeking behavior in rats. *J. Neurosci.*, *20*, 7489–7495.

Ito, R., Dalley, J. W., Robbins, T. W., & Everitt, B. J. (2002). Dopamine release in the dorsal striatum during cocaine-seeking behavior under the control of a drug-associated cue. *J. Neurosci.*, *22*, 6247–6253.

Johnsrude, I. S., Owen, A. M., White, N. M., Zhao, W. V., & Bohbot, V. (2000). Impaired preference conditioning after anterior temporal lobe resection in humans. *J. Neurosci.*, *20*, 2649–2656.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.*, *10*, 1625–1633.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *Am. Psychol.*, *58*, 697–720.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kalivas, P. W., & Duffy, P. (1995). Selective activation of dopamine transmission in the shell of the nucleus accumbens by stress. *Brain Res.*, *675*, 325–328.

Kawagoe, R., Takikawa, Y., & Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.*, *1*, 411–416.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, *308*, 78–83.

Kirsch, P., Schienle, A., Stark, R., Sammer, G., Blecker, C., Walter, B., Ott, U., Burkart, J., & Vaitl, D. (2003). Anticipation of reward in a nonaversive differential conditioning paradigm and the brain reward system: An event-related fMRI study. *NeuroImage*, *20*, 1086–1095.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J. Neurosci.*, *21*, RC159.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *J. Neurosci., 25*, 4806–4812.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*, 673–676.

Kötter, R., & Meyer, N. (1992). The limbic system: A review of its empirical foundation. *Behav. Brain Res.*, *52*, 105–127.

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fMRI study. *Neuron*, *20*, 937–945.

LaBar, K. S., LeDoux, J. E., Spencer, D. D., & Phelps, E. A. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *J. Neurosci.*, *15*, 6846–6855.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.*, *23*, 155–184.

LeDoux, J. E., & Gorman, J. M. (2001). A call to action: Overcoming anxiety through active coping. *Am. J. Psychiatry*, *158*, 1953–1955.

Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart strings and purse strings: Effects of emotions on economic transactions. *Psychol. Sci.*, *15*, 337–341.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150–157.

Lynd Balta, E., & Haber, S. N. (1994). The organization of midbrain projections to the striatum in the primate: Sensorimotor-related striatum versus ventral striatum. *Neuroscience*, *59*, 625–640.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*, 339–346.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503–507.

McCullough, L. D., Sokolowski, J. D., & Salamone, J. D. (1993). A neurochemical and behavioral investigation of the involvement of nucleus accumbens dopamine in instrumental avoidance. *Neuroscience*, *52*, 919–925.

McGaugh, J. L. (2000). Memory—A century of consolidation. *Science, 287*, 248–251.

Middleton, F. A., & Strick, P. L. (1997). New concepts about the organization of basal ganglia output. *Adv. Neurol.*, *74*, 57–68.

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, *36*, 265–284.

Murphy, C. A., Pezze, M., Feldon, J., & Heidbreder, C. (2000). Differential involvement of dopamine in the shell and core of the nucleus accumbens in the expression of latent inhibition to an aversively conditioned stimulus. *Neuroscience*, *97*, 469–477.

Nieuwenhuis, S., Heslenfeld, D. J., von Geusau, N. J., Mars, R. B., Holroyd, C. B., & Yeung, N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. *NeuroImage*, *25*, 1302–1309.

Ochsner, K. N. (2004). Current directions in social cognitive neuroscience. *Curr. Opin. Neurobiol.*, *14*, 254–258.

Ochsner, K. N., & Gross, J. J. (2008). Cognitive emotion regulation: Insights from social, cognitive and affective neuroscience. *Curr. Dir. Psychol. Sci.*, *17*, 153–158.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.*, *14*, 769–776.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*, 329–337.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*, 452–454.

O'Doherty, J. P., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron*, *33*, 815–826.

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fears by observing others: The neural systems of social fear transmission. *Soc., Cogn. Affect. Neurosci.*, *2*, 2–10.

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nat. Neurosci.*, *10*, 1095–1102.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.*, *25*, 563–593.

Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nat. Neurosci.*, *5*, 97–98.

Parkinson, J. A., Crofts, H. S., McGuigan, M., Tomic, D. L., Everitt, B. J., & Roberts, A. C. (2001). The role of the primate amygdala in conditioned reinforcement. *J. Neurosci.*, *21*, 7770–7780.

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annu. Rev. Psychol.*, *57*, 27–53.

Phelps, E. A. (2008). The study of emotion in neuroeconomics. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics* (pp. 233–250). New York: Elsevier.

Phelps, E. A., Delgado, M. R., Nearing, K. I., & LeDoux, J. E. (2004). Extinction learning in humans: Role of the amygdala and vmPFC. *Neuron*, *43*, 897–905.

Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C., & Davis, M. (2001). Activation of the left amygdala to a cognitive representation of fear. *Nat. Neurosci.*, *4*, 437–441.

Ploghaus, A., Tracey, I., Clare, S., Gati, J. S., Rawlins, J. N., & Matthews, P. M. (2000). Learning about pain: The neural substrate of the prediction error for aversive events. *Proc. Natl. Acad. Sci. USA*, *97*, 9281–9286.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*, 546–550.

Porcelli, J., & Delgado M. R. (2009). Acute stress modulates risk taking in financial decision making. *Psychol. Sci., 20*(3), 278–283.

Robbins, T. W., & Everitt, B. J. (1992). Functions of dopamine in the dorsal and ventral striatum. *Sem. Neurosci.*, *4*, 119–127.

Robinson, T. E., Becker, J. B., Young, E. A., Akil, H., & Castaneda, E. (1987). The effects of footshock stress on regional brain dopamine metabolism and pituitary beta-endorphin release in rats previously sensitized to amphetamine. *Neuropharmacology*, *26*, 679–691.

Rosenkranz, J. A., & Grace, A. A. (2002). Dopamine-mediated modulation of odour-evoked amygdala potentials during Pavlovian conditioning. *Nature*, *417*, 282–287.

Rosenkranz, J. A., Moore, H., & Grace, A. A. (2003). The prefrontal cortex regulates lateral amygdala neuronal plasticity and responses to previously conditioned stimuli. *J. Neurosci.*, *23*, 11054–11064.

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science*, *310*, 1337–1340.

Schacter, S., & Singer J. (1962). Cognitive, social and physiological determinants of emotional state. *Psychol. Rev.*, *29*, 379–399.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Soc. Sci. Inf.*, *44*, 695–729.

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J. Neurosci.*, *27*, 12860—12867.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annu. Rev. Neurosci.*, *23*, 473–500.

Schultz, W., Preuschoff, K., Camerer, C., Hsu, M., Fiorillo, C. D., Tobler, P. N., & Bossaerts, P. (2008). Review. Explicit neural signals reflecting reward uncertainty. *Philos. Trans. R Soc. Lond. B Biol. Sci.*, *363*, 3801–3811.

Schwarting, R., & Carey, R. J. (1985). Deficits in inhibitory avoidance after neurotoxic lesions of the ventral striatum are neurochemically and behaviorally selective. *Behav. Brain Res.*, *18*, 279–283.

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *J. Neurosci.*, *25*, 2941–2951.

Setlow, B., Holland, P. C., & Gallagher, M. (2002). Disconnection of the basolateral amygdala complex and nucleus accumbens impairs appetitive Pavlovian second-order conditioned responses. *Behav. Neurosci.*, *116*, 267–275.

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *J. Neurosci.*, *27*, 4826–4831.

Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nat. Neurosci. 8*, 1234–1240.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron*, *41*, 653–662.

Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., & Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *Proc. Natl. Acad. Sci. USA*, *106*, 5035–5040.

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2007). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.*, *97*, 1621–1632.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*, 515–518.

Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron*, *41*, 281–292.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nat. Neurosci.*, *7*, 1271–1278.

Voorn, P., Vanderschuren, L. J. M. J., Groenewegen, H. J., Robbins, T. W., & Pennartz, C. M. A. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends Neurosci., 27*, 468–474.

Wadenberg, M. L., Ericson, E., Magnusson, O., & Ahlenius, S. (1990). Suppression of conditioned avoidance behavior by the local application of (–)sulpiride into the ventral, but not the dorsal, striatum of the rat. *Biol. Psychiatry*, *28*, 297–307.

White, N. M., & Salinas, J. A. (2003). Mnemonic functions of dorsal striatum and hippocampus in aversive conditioning. *Behav. Brain Res.*, *142*, 99–107.

Winocur, G., & Mills, J. A. (1969). Effects of caudate lesions on avoidance behavior in rats. *J. Comp. Physiol. Psychol.*, *68*, 552–557.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.*, *7*, 464–476.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2005). Blockade of NMDA receptors in the dorsomedial striatum abolishes action-outcome learning in instrumental conditioning. *Eur. J. Neurosci., 22*, 505–512.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci., 22*, 513–523.

Young, A. M. (2004). Increased extracellular dopamine in nucleus accumbens in response to unconditioned and conditioned aversive stimuli: studies using 1 min microdialysis in rats. *J. Neurosci. Methods*, *138*, 57–63.

Young, M. P. (2002). Connextional organisation of the function in the macaque cerebral cortex. In A. Schuz & R. Miller (Eds.), *Cortical areas: Unity and diversity* (pp. 355–376). New York: CRC Press.

Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., & Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, *42*, 509–517.

X

CONSCIOUSNESS

Chapter 77 BLOCK 1111

78 SCHIFF 1123

79 KOCH 1137

80 REES 1151

81 MACKNIK AND MARTINEZ-CONDE 1165

82 KOENIGS AND ADOLPHS 1181

83 LAU 1191

84 TONONI AND BALDUZZI 1201

Introduction

CHRISTOF KOCH

THE NATURE OF the relationship between the objective, material world and the subjective world of conscious experience, of phenomenal content, lies at the heart of the ancient mind-body problem. Consciousness appears as mysterious to 21st century scholars as it was to the ancient Greeks who first systematically contemplated the nature of mind. Yet modern science is in a strong position to investigate and manipulate the physical basis of consciousness. Brain scientists seek an understanding of how and why the neural basis of one particular conscious sensation is the basis of that sensation rather than another, why so many behaviors occur outside the pale of consciousness, why sensations are structured the way they are, and how they acquire meaning. Finally, cognitive neuroscience is contributing to the problem of understanding the conscious sensation of willing an action.

The scholars represented in section X take the problem of consciousness, the first-person perspective, as given and assume that brain activity is both necessary and sufficient for biological creatures to experience something. A primary goal is to identify the specific nature of the activity of brain cells that gives rise to any one specific conscious percept, the neuronal correlates of consciousness (Crick & Koch, 1995; Metzinger, 2000; Chalmers, 2000). An auxiliary goal is to determine to what extent these correlates differ from activity that influences behavior without engaging consciousness.

Most everyone has a general idea of what it means to be conscious. According to the philosopher John Searle, "Consciousness consists of those states of sentience, or feeling, or awareness, which begin in the morning when we awake from a dreamless sleep and continue throughout the day until we fall into a coma or die or fall asleep again or otherwise become unconscious" (Searle, 1997). Practically speaking, consciousness is needed for nonroutine tasks that require retention of information over seconds. Although

provisional and vague, such a definition is good enough to get the process started. As the science of consciousness advances, such operational definitions will need to be refined and expressed in more fundamental neuronal terms. Until the problem is understood much better, though, a more formal definition is likely to be either misleading or overly restrictive, or both. If this statement seems evasive, try defining a gene (Keller, 2000).

The working hypothesis is that consciousness emerges from neuronal features of the brain. Emergence is used here without any new-age overtones but in the sense that the initiation and propagation of the action potential in axons, a highly nonlinear phenomenon, is the result of—and can be predicted from—the attributes of voltage-dependent ionic channels inserted into the neuronal membrane. Although consciousness is fully compatible with the laws of physics, it is not obvious how it follows from these laws. Something else might be needed.

Understanding the material basis of consciousness is unlikely to require any exotic new physics but rather a much deeper appreciation of how highly interconnected networks of a very large number of heterogeneous but highly structured neurons work. The abilities of rapidly forming and dissolving coalitions of corticothalamic neurons to learn from interactions with the environment and from their own internal activities are routinely underestimated. The individual neurons themselves are complex entities with unique morphologies and thousands of inputs and outputs. Humans have no real experience with such vast organization. Hence, even biologists struggle to appreciate the properties and power of the nervous system.

It would be contrary to evolutionary continuity to believe that consciousness is unique to humans. Most brain scientists assume that many species possess some, but not necessarily all, of the features of consciousness—that they see, hear, smell, and otherwise experience the world (Griffin, 2001). This assumption is particularly true for monkeys and apes, whose behavior, development, and brain structure are remarkably similar to those of humans. Of course, each species has its own unique sensorium, matched to its ecological niche, but that is not to deny that animals can have feeling, subjective states. To believe otherwise seems presumptuous and flies in the face of all experimental evidence. At this point in time, we have no clear idea whether animals from phyla other than Chordata have subjective states. But the nonstereotyped and adaptive behaviors of many mollusks and insects make it likely that they too share the gift of consciousness with us.

The focus of much of the empirical work in the field and in this section of the book is on visual consciousness. More than other aspects of sensation, visual awareness is amenable to empirical investigations for a variety of reasons. First, humans are visual creatures. This fact is reflected in the large amount of brain tissue dedicated to vision and in the importance of seeing in daily life. Second, images are highly structured yet easy to control using computers. Third, phenomena such as binocular rivalry, continuous flash suppression, or motion-induced blindness can be used to manipulate the relationship between retinal input and visibility—that is, between objective sensory stimulus and subjective conscious percept. Last, the neuronal basis of many visual phenomena and illusions has been investigated throughout the animal kingdom. Perceptual neuroscience has advanced to such a point that reasonably sophisticated computational models have been constructed and have proven their worth in guiding experimental agendas and summarizing the data. It is not unlikely that all the different aspects of consciousness (smell, pain, vision, self-consciousness, the feeling of willing an action, and so on) employ one or perhaps a few common mechanisms. Figuring out the neuronal basis for one modality, therefore, will probably be the breakthrough event that will help us understand all of them.

Much has changed in consciousness studies in the five years since the previous edition of this book. So much so that an almost entirely new cast of philosophers and scientists discuss their findings in these pages, the vast majority of which have been published within the past few years. Better than anything else, this demonstrates the vitality of this research endeavor.

With the preceding background, let me briefly introduce the eight chapters. In chapter 77, Block summarizes and compares the three major theoretical approaches to consciousness that take science seriously: higher order theories, the global workspace account of consciousness, and biological theories—those that postulate that consciousness is some sort of biological state of the brain. Block, a philosopher, has the distinction of having himself directly contributed to the empirical debate, by drawing subtle but crucial distinctions between selective attention and consciousness and their underlying neuronal mechanisms (Block, 2007). In chapter 78, Schiff discusses the clinical literature pertaining to global impairments of consciousness following brain injury. The persistent vegetative state (PVS) or the minimal conscious state (MCS) are clinical conditions in which the patient is either permanently unconscious or hovering at the borderline between unconscious and conscious and in which midline structures in the brain stem and thalamus are affected. Given the large number of such patients—on the order of 100,000 in the United States alone—there is a great urgency to understand these pathologies. The exploratory work of Schiff and colleagues (2007) using thalamic deep-brain stimulation is particularly promising in this regard. Chapter 79 by Koch provides the conceptual framework and empirical data for a research program dedicated to discovering the neuronal basis of the content of conscious perception. It emphasizes what is not needed for consciousness

(e.g., neither sensory input nor motor output nor self-consciousness), the dissociation between attention and consciousness, and the interaction among coalitions of corticothalamic neurons that vie for dominance. Chapter 80 by Rees and chapter 81 by Macknik and Martinez-Conde focus on different aspects of vision. Rees describes the enormous contributions functional brain imaging has made to elucidate the cortical basis of visual awareness in both patients and neurologically normal people; in contrast, Macknik and Martinez-Conde concentrate on the anatomy and electrophysiology of the early stages of the visual thalamocortical system in monkeys and what they teach us about consciousness. One of the crucial conclusions is that attention and consciousness, so often conflated, are quite distinct processes. Koenigs and Adolphs' chapter 82 deals with data and theories that seek to explain the neuronal basis of the subjective states of joy, sorrow, anger, and fear. They rightfully emphasize that much emotional processing can occur outside the pale of consciousness. While most students of the mind have gotten used to unconscious visual processing, the idea of subliminal emotional processing is not (yet) as widely accepted. Lau's chapter 83 deals with the problem of the function of consciousness in the context of willed actions. What is not in doubt is the fact that most of us consciously experience a feeling of willing an action, such as raising our arm. What is more controversial is whether this feeling has any causal effect on the behavior. Is the sensation of volition simply an epiphenomenon? Lau challenges the community to come up with experiments that conclusively demonstrate that some function cannot be carried out unconsciously. It is not an easy challenge to meet. Last, the field of consciousness studies is greatly hampered by the lack of a widely accepted theory of consciousness. We need to know, on theoretical or conceptual grounds, why a particular system of interacting parts has subjective states. Is our immune system conscious? And if not, why not? What about a fetus, a newborn child, an aphasic patient, a monkey, or a honeybee? And can we replicate consciousness in silicon? These questions are addressed in chapter 84 by Tononi and Balduzzi. Proceeding from two simple phenomenological axiomatic observations, they construct an integrated theory of consciousness, including qualia space, that explains many of the known anatomical, physiological, and psychological facts about consciousness. It is a very ambitious but enormously exciting development.

Collectively, the chapters in section X signal the emergence of a science of consciousness, an ability to investigate how phenomenal feelings emerge out of excitable brain matter in a rigorous, reliable, and reproducible manner. What are needed now are invasive experiments that begin to close the gap between correlation and causation. Molecular biology is delivering optogenetic techniques to deliberately, delicately, transiently, and reversibly dissect individual components of forebrain circuits in flies, mice, and, soon, monkeys (Adamantidis, Zhang, Aravanis, Deisseroth, & de Lecea, 2007). The applications of such technologies, in combination with continuous, long-term recordings from thousands of neurons and functional imaging techniques, will do much to advance this goal.

REFERENCES

Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K., & de Lecea, L. (2007). Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature*, *450*, 420–424.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.*, *30*, 481–548.

Chalmers, D. J. (2000) What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions* (pp. 17–40). Cambridge, MA: MIT Press.

Crick, F. C., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, *375*, 121–123.

Griffin, D. R. (2001). *Animal minds: Beyond cognition to consciousness.* Chicago: University of Chicago Press.

Keller, E. F. (2000). *The century of the gene.* Cambridge, MA: Harvard University Press.

Metzinger, T. (Ed.). (2000). *Neural correlates of consciousness: Empirical and conceptual questions*. Cambridge, MA: MIT Press.

Schiff, N. D., Giacino, J. T., Kalmar, K., Victor, J. D., Baker, K., Gerber, M., Fritz, B., Eisenberg, B., O'Connor, J., Kobylarz, E. J., Farris, S., Machado, A., McCagg, C., Plum, F., Fins, J. J., & Rezai, A. R. (2007). Behavioral improvements with thalamic stimulation after severe traumatic brain injury. *Nature*, *448*, 600–603.

Searle, J. R. (1997). *The mystery of consciousness.* New York: New York Review of Books.

77 Comparing the Major Theories of Consciousness

NED BLOCK

ABSTRACT This article compares the three frameworks for theories of consciousness that are taken most seriously by neuroscientists: the view that consciousness is a biological state of the brain, the global workspace perspective, and an account in terms of higher order states. The comparison features the "explanatory gap" (Nagel, 1974; Levine, 1983), the fact that we have no idea why the neural basis of an experience is the neural basis of that experience rather than another experience or no experience at all. It is argued that the biological framework handles the explanatory gap better than do the global workspace or higher order views. The article does not discuss quantum theories or "panpsychist" accounts according to which consciousness is a feature of the smallest particles of inorganic matter (Chalmers, 1996; Rosenberg, 2004). Nor does it discuss the "representationist" proposals (Tye, 2000; Byrne, 2001a) that are popular among philosophers but not neuroscientists.

Three theories of consciousness

HIGHER ORDER The higher order approach says that an experience is phenomenally conscious only in virtue of *another* state that is about the experience (Armstrong, 1978; Lycan, 1996a; Byrne, 1997; Carruthers, 2000; Byrne, 2001b; Rosenthal, 2005a). This perspective comes in many varieties, depending on, among other things, whether the monitoring state is a thought or a perception. The version to be discussed here says that the higher order state is a thought ("higher order thought" is abbreviated as HOT) and that a conscious experience of red consists in a representation of red in the visual system accompanied by a thought in the same subject to the effect that the subject is having the experience of red.

GLOBAL WORKSPACE The global workspace account of consciousness was first suggested by Bernard Baars (1988) and has been developed in a more neural direction by Stanislas Dehaene, Jean-Pierre Changeux, and their colleagues (Dehaene, Changeux, Nacchache, Sackur, & Sergent, 2006). The account presupposes a neural network approach in which there is competition among neural coalitions involving both frontal and sensory areas (Koch, 2004), the winning coalitions being conscious. Sensory stimulation causes activations in sensory areas in the back of the head that compete with each other to form dominant coalitions (indicated by dark elements in the outer rings in figure 77.1). Some of these dominant coalitions trigger central reverberations through long-range connections to frontal cortex, setting up activations that help to maintain both the central and peripheral activations. The idea that some brain areas control activations and reactivations in other areas is now ubiquitous in neuroscience (Damasio & Meyer, 2008), and a related idea is widely accepted: that one instance of reciprocal control is one in which workspace networks in frontal areas control activations in sensory and spatial areas (Curtis & D'Esposito, 2003). It is useful in thinking about the account to distinguish between suppliers and consumers of representations. Perceptual systems supply representations that are consumed by mechanisms of reporting, reasoning, evaluating, deciding, and remembering, which themselves produce representations that are further consumed by the same set of mechanisms. Once perceptual information is "globally broadcast" in frontal cortex this way, it is available to all cognitive mechanisms without further processing. Phenomenal consciousness is global broadcasting.

Although the global workspace account is motivated and described in part in neural terms, the substantive claims of the model abstract away from neuronal details. Nothing in the model requires the electrochemical nature of actual neural signals. The architectural aspects of the model can just as easily be realized in silicon-based computers as in protoplasm. In this respect, the global workspace theory of consciousness is a form of what philosophers call *functionalism* (Block, 1980), according to which consciousness is characterized by an abstract structure that does not include the messy details of neuroscience.

Another functionalist theory of consciousness is the integrated information theory (Tononi & Edelman, 1998), according to which the level of consciousness of a system at a time is a matter of how many possible states it has at that time and how tightly integrated its states are. This theory has a number of useful features—for example, retrodicting that there would be a loss of consciousness in a seizure in which the number of possible states drops precipitously

NED BLOCK Departments of Philosophy and Psychology; Center for Neural Science, New York University, New York, New York

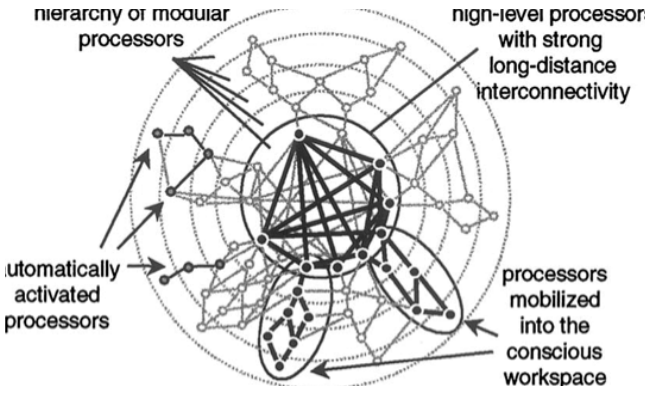


FIGURE 77.1 Schematic diagram of the global workspace. Sensory activations in the back of the brain are symbolized by dots and lines in the outside rings. Dominant sensory neural coalitions (dark lines and dots) compete with one another to trigger reverberatory activity in the global workspace (located in frontal areas) in the center of the diagram. The reverberatory activity in turn maintains the peripheral excitation until a new dominant coalition wins out.

(Tononi & Koch, 2008). Unfortunately, such predictions would equally follow from an integrated information theory of *intelligence* (in the sense of the capacity for thought, as in the Turing test of intelligence)—which also drops in a seizure. Consciousness and intelligence are on the face of it very different things. We all understand science fiction stories in which intelligent machines lack some or all forms of consciousness. And on the face of it, mice or even lower animals might have phenomenal consciousness without much intelligence. The separation of consciousness and cognition has been crucial to the success of the scientific study of consciousness. In a series of papers that established the modern study of consciousness, Crick and Koch (1990, 1998) noted in particular that the basic processes of visual consciousness could be found in nonprimate mammals and were likely to be independent of language and cognition. Although its failure to distinguish consciousness and intelligence is crippling for the current prospects of the integrated information theory as a stand-alone theory of consciousness, I will mention it at the end of the article in a different role: as an adjunct to a biological theory.

THE BIOLOGICAL THEORY The third of the major theories is the biological theory, the theory that consciousness is some sort of biological state of the brain. It derives from Democritus (Kirk, Raven, & Schofield, 1983) and Hobbes (1989), but was put in modern form in the 1950s by Place (1956), Smart (1959), and Feigl (1958). (See also Block, 1978; Crane, 2000; Lamme, 2003.) I will explain it using as an example the identification of the visual experience of (a kind of) motion in terms of a brain state that includes activations of a certain sort of area MT+ in the visual cortex. Although this explanation is useful as an example, we can expect that any theory of visual experience will be superseded.

Visual area MT+ reacts to motion in the world, different cells reacting to different directions. Damage to MT+ can cause loss of the capacity to experience this kind of motion; MT+ is activated by the motion aftereffect; transcranial magnetic stimulation of MT+ disrupts these afterimages and also can cause motion "phosphenes" (Zihl, von Cramon, & Mai, 1983; Britten, Shadlen, Newsome, & Movshon, 1992; Heeger, Boynton, Demb, Seideman, & Newsome, 1999; Kammer, 1999; Cowey & Walsh, 2000; Kourtzi & Kanwisher, 2000; Huk, Ress, & Heeger, 2001; Rees, Kreiman, & Koch, 2002; Théoret, Kobayashi, Ganis, Di Capua, & Pascual-Leone, 2002). However, it is important to distinguish between two kinds of MT+ activations, which I will call nonrepresentational activations and representational activations. Some activations in the visual system are very weak, do not "prime" other judgments (that is, do not facilitate judgments about related stimuli), and do not yield above-chance performance on forced-choice identification or detection (that is, they do not allow subjects to perform above chance on a choice of what the stimulus was or even whether there was a stimulus or not). On a very liberal use of the term "representation" in which *any* neural activation that correlates with an external property is a representation of it (Gallistel, 1998), one might nonetheless call such activations of MT+ representations, but it will be useful to be less liberal here, describing the weak activations just mentioned as nonrepresentational. (The term "representation" is very vague and can be made precise in different equally good ways.) However, if activations of MT+ are strong enough to be harnessed in subjects' choices (at a minimum in priming), then we have genuine representations. (See Siegel, 2008, for a discussion of the representational contents of perceptual states.)

Further, there is reason to think that representations in MT+ that also generate feedback loops to lower areas are at least potentially conscious representational contents (Pascual-Leone & Walsh, 2001; Silvanto, Cowey, Lavie, & Walsh, 2005). (For a dissident anti-feedback-loop perspective, see Macknik & Martinez-Conde, 2007.) Of course, an activated MT+ even with feedback to lower visual areas is not all by itself sufficient for phenomenal consciousness. No one thinks that a section of visual cortex in a bottle would be conscious (Kanwisher, 2001).

What makes such a representational content phenomenally conscious? One suggestion is that active connections between cortical activations and the top of the brain stem constitute what Alkire, Haier, and Fallon (2000) call a "thalamic switch." There are two important sources of evidence for this view. One is that the common feature of many if not all anesthetics appears to be that they disable these connections (Alkire & Miller, 2005). Another is that the transition from the vegetative state to the minimally conscious state (Laureys, 2005) involves these connections. However, there

is some evidence that the "thalamic switch" is an on switch rather than an off switch (Alkire & Miller, 2005) and that corticothalamic connections are disabled as a result of the large overall decrease in cortical metabolism (Velly et al., 2007; Alkire, 2008; Tononi & Koch, 2008)—which itself may be caused in part by the deactivation of other subcortical structures (Schneider & Kochs, 2007). Although this area of study is in flux, the important philosophical point is the three-way distinction between (1) a nonrepresentational activation of MT+, (2) an activation of MT+ that is a genuine visual representation of motion, and (3) an activation of MT+ that is a key part of a phenomenally conscious representation of motion.

The same distinctions can be seen in terms of the global workspace theory as the distinction among (1) a minimal sensory activation (the gray peripheral activations in figure 77.1), (2) a peripheral dominant coalition (the black peripheral activations in figure 77.1), and (3) a global activation involving both peripheral and central activation (the circled activations in figure 77.1 that connect to the central workspace). The higher order account is focused on the distinction between a visual representation and a conscious visual representation (2 versus 3), a visual representation that is accompanied by a higher order thought to the effect that the subject has it.

Here are some items of comparison between the three theories. According to the biological account, global broadcasting and higher order thought are what consciousness *does* rather than what consciousness *is*. That is, one function of consciousness on the biological view is to promote global broadcasting, and global broadcasting in some but not all cases can lead to higher order thought. Further, according to the biological view, both the global workspace and higher-order thought views leave out too many details of the actual working of the brain to be adequate theories of consciousness. Information in the brain is coded electrically, then transformed to a chemical code, then back to an electrical code, and it would be foolish to assume that this transformation from one form to another is irrelevant to the physical basis of consciousness.

From the point of view of the biological and global workspace views, the higher-order-thought view sees consciousness as more intellectual than it is, but from the point of view of higher-order-thought accounts, the biological and global workspace accounts underestimate the role of cognition in consciousness. The global workspace and higher-order-thought accounts are sometimes viewed as superior to the biological account in that the biological account allows for the possibility that a subject could have a phenomenally conscious state that the subject does not know about (Block, 2007a, 2007b). And this is connected to the charge that the biological account—as compared with the other accounts—neglects the connection between phenomenal consciousness and the self (Church, 1995; Harman, 1995; Kitcher, 1995).

The higher order and global workspace accounts link consciousness to the ability to report it more tightly than does the biological view. On the higher-order-thought view, reporting is just expressing the higher order thought that makes the state conscious, so the underlying basis of the ability to report comes with consciousness itself. On the global workspace account, what makes a representational content conscious is that it is in the workspace, and that just is what underlies reporting. On the biological account, by comparison, the biological machinery of consciousness has no necessary relation to the biological machinery underlying reporting, and hence there is a real empirical difference among the views that each side seems to think favors its own view (Block, 2007b; Naccache & Dehaene, 2007; Prinz, 2007; Sergent & Rees, 2007).

To evaluate and further compare the theories, it will be useful to appeal to a prominent feature of consciousness, the explanatory gap.

The explanatory gap

Phenomenal consciousness is "what it is like" to have an experience (Nagel, 1974). Any discussion of the physical basis of phenomenal consciousness (henceforth just consciousness) has to acknowledge the "explanatory gap" (Nagel, 1974; Levine, 1983): nothing that we now know, indeed nothing that we have been able to hypothesize or even fantasize, gives us an understanding of why the neural basis of the experience of green that I now have when I look at my screen saver is the neural basis of *that* experience as opposed to *another* experience or no experience at all. Nagel puts the point in terms of the distinction between subjectivity and objectivity: the experience of green is a subjective state, but brain states are objective, and we do not understand how a subjective state could *be* an objective state or even how a subjective state could be *based in* an objective state. The problem of closing the explanatory gap (the "Hard Problem" as Chalmers, 1996, calls it) has four important aspects: (1) we do not see a hint of a solution; (2) we have no good argument that there is no solution that another kind of being could grasp or that we may be able to grasp at a later date (but see McGinn, 1991); so (3) the explanatory gap is not intrinsic to consciousness; and (4) most importantly for current purposes, recognizing the first three points requires no special theory of consciousness. All scientifically oriented accounts should agree that consciousness is in some sense based in the brain; once this fact is accepted, the problem arises of why the brain basis of this experience is the basis of this one rather than another one or none, and it becomes obvious that nothing now known gives a hint of an explanation.

The explanatory gap was first brought to the attention of scientists through the work of Nagel (1974) and Crick and Koch (Crick, 1994; Crick & Koch, 1998). Many would argue that the candid recognition of what we do not understand played an important role in fueling the incredible wave of research that still engulfs us.

How do the three theories account for the explanatory gap?

The HOT view says that consciousness of, say, red is a matter of three ingredients: a higher order thought, a representation with the content red, and an aboutness relation between the first and the second. According to the HOT perspective, each of these ingredients can exist individually without any consciousness. We have unconscious thoughts—for example, subliminal representations of red—and those unconscious thoughts are, unconsciously, about things. According to the HOT theory, if a subject has an unconscious representation of red, and then forms an unconscious thought about the representation of red, the representation of red automatically is conscious. Of course, in some trivial sense of "conscious" we might decide to *call* that representation of red conscious, meaning *only* that there is a higher-order thought about it; but if the HOT theory is about consciousness in the full-blooded sense in which for a state to be conscious is for there to be something it is like to be in that state, there is a fundamental mystery for the HOT view.

It may seem that this is just *the* explanatory gap in a new form, one appropriate to the HOT theory, but that assertion is a mistake. Consider the prime order thought (POT) view—which says that thoughts about thoughts about thoughts . . . are always conscious so long as the number of embeddings is prime. There is a puzzle of the *POT view's own making* of why a prime number of embeddings creates consciousness, but that puzzle is not the *real* explanatory gap.

The real explanatory gap is the problem of why the neural basis of a conscious state with a specific conscious quality is the neural basis of that conscious quality rather than another or nothing at all. The real explanatory gap does not assume any specific theory except the common basis of all scientific approaches in the 21st century, that conscious qualities *have* a brain basis.

The problem for the HOT perspective is that it is part of the idea of it that putting together ingredients that are not in themselves conscious (thought, aboutness, and representation) automatically exhibits consciousness. The most neuroscience can do is explain thought, explain aboutness, and explain representation. But there is no reason to expect—and it is not part of any HOT perspective—that neuroscience will find some magic glow that occurs when those things combine.

The fact that the HOT theory cannot recognize the real explanatory gap makes it attractive to people who do not agree that there is an explanatory gap in the first place—the HOT theory is a kind of "no consciousness" theory of consciousness. But for those who accept an explanatory gap (at least for our current state of neuroscientific knowledge), the fact that the HOT theory does not recognize one is a reason to reject the HOT theory. The HOT theory is geared to the cognitive and representational aspect of consciousness, but if those aspects are not the whole story, the HOT theory will never be adequate to consciousness.

This very short argument against the HOT approach also applies to the global workspace theory, albeit in a slightly different form. According to the global workspace account, the answer to the question of why the neural basis of my experience of red is the neural basis of a conscious experience is simply that it is globally broadcast. But why is a globally broadcast representation conscious? This is indeed a puzzle for the global workspace theory but it is not the explanatory gap because it presupposes the global workspace theory itself, whereas the explanatory gap (discussed previously) does not.

The most neuroscience can do for us according to the global workspace account is explain how a representation can be broadcast in the global workspace, but the task will still remain of explaining why global broadcasting, however realized, is conscious. In principle, global broadcasting could be realized in an electronic system rather than a biological system, and of course the same issue will arise. So that issue cannot be special to the biological realization of mind.

The biological account, by contrast, fits the explanatory gap—indeed, I phrased the explanatory gap in terms of the biological account, asking how we can possibly understand how consciousness could be a biological property. So the biological account is the only one of the three major theories to fully acknowledge the explanatory gap. From the point of view of the HOT and global workspace theories, their task concerning the explanatory gap is not to show how they can accommodate it, but rather to *explain away our impression that there is one*. One such attempt will be considered in the next section.

There is a fine line between acknowledging the explanatory gap and surrendering to dualism, as also discussed in the next section.

The explanatory gap and dualism

Dualism is the view that there is some aspect of the mind that is not physical (Chalmers, 1996). It comes in many varieties, but the issues to be discussed do not depend on any specific variety.

Let us start with a historical analogy (Nagel, 1974). A pre-Socratic philosopher would have no way of understanding how heat could be a kind of motion or of how light could be a kind of vibration. Why? Because the pre-Socratic

philosopher did not have the appropriate concepts of motion—namely, the concept of kinetic energy and its role—or of vibration—namely, the concepts involved in the wave theory of light—that would allow an understanding of how such different concepts could pick out the same phenomenon.

What is a concept? A concept is a mental representation usable in thought. We often have more than one concept of the same thing. The concept *light* and the concept *electromagnetic radiation of 400–700 nm* pick out the same phenomenon. What the pre-Socratic philosopher lacks is a concept of light and an appropriate concept of vibration (one that requires a whole theory). What is missing for the pre-Socratic is not just the absence of a theoretical definition but a lack of understanding of what things are *grouped together* from a scientific point of view. We now realize that ripples in a pond, sound, and light are all phenomena of the same kind: waves. And we now realize that burning, rusting, and metabolizing are all cases of oxidation (Churchland, 2002), but the pre-Socratics, given their framework in which the basic categories were fire, earth, air, and water, would have had no way to grasp these facts. One upshot is that if superscientists of the future were to tell us what consciousness is, we probably would not have the conceptual machinery to understand, just as the pre-Socratic would not have the conceptual machinery to understand that heat is a kind of motion or that light is a kind of vibration.

Armed with this idea, we can see how to steer between the explanatory gap and dualism. What we lack is an objective neuroscientific concept that would allow us to see how it could pick out the same phenomenon as our subjective concept of the experience of green. And we can expect that we do not even have the right subjective concept of the experience of green, since we are not sure what subjective phenomena truly should be grouped together. The resolution of the apparent conflict between the explanatory gap and physicalism is that subjectivity and objectivity can be seen as properties of *concepts* rather than properties of the states that the concepts are concepts of. This idea, that we can see arguments that apparently indicate ontological dualism—that is, a dualism of objects or substances or properties—as really an argument for *conceptual* dualism, stems from Nagel (1974) and Loar (1990/1997) and is sometimes called New Wave physicalism (see Horgan & Tienson, 2001).

Another way of seeing the point is to consider Jackson's (1982) famous thought experiment concerning Mary, a neuroscientist of the distant future who knows everything there is to know about the scientific basis of color experience, but has grown up in a black-and-white environment. When she sees red for the first time, she learns what it is like to see red, despite already knowing all the scientific facts about seeing red. Does this show that the fact of what it is like to see red is not a scientific fact? No, because we can think of what Mary learns in terms of her acquiring a subjective *concept* of a state that *she already had an objective concept of*. Imagine someone who already knows that Lake Michigan is filled with H_2O, but learns something new: that Lake Michigan is filled with water. What this person learns is not a new fact but a new piece of knowledge, involving a new concept, of a fact the person already knew. Similarly, Mary acquires new knowledge, but that new knowledge does not go beyond the scientific facts that she already knew about, and so does not support any kind of dualism. (This line of thought is debated in Block, 2006; White, 2006.)

Importantly, this line of reasoning *does not do away with the explanatory gap* but rather reconceives it as a failure to understand how a subjective and an objective concept can pick out the same thing.

These points about different concepts of the same thing have sometimes been used to try to dissolve the explanatory gap (Papineau, 2002). The idea is that the false appearance of an explanatory gap arises from the gap between a subjective concept of a phenomenally conscious state and an objective concept of the same state. But note: I can think the thought that the color I am now experiencing as I look at an orange (invoking a subjective concept of orange) is identical to the color between red and yellow (invoking an objective concept of orange). But this use of the two kinds of concepts engenders no explanatory gap.

Thus far, the score is biological theory 1, HOT and global workspace 0. But the competition has not yet encountered the heartland of the HOT theory.

Consciousness-of

It is very often (but not always—Dretske, 1993) assumed that a conscious state is a state that one is conscious of being in (Lycan, 1996a). I am willing to agree in order to focus on other matters. The HOT theory has an attractive explanation of this claim, because consciousness-of can be cashed out as being the object of a HOT. However, there are two other accounts of why a conscious state is one that one is conscious of being in, and these accounts are preferable to the HOT account—according to the viewpoint of the biological theory and the global workspace theory. The deflationary account (Sosa, 2003) says that all there is to being conscious of one's experience is the triviality that in having an experience, one experiences it, just as one smiles one's smile and dances one's dance. Consciousness-of in this sense is to be firmly distinguished from attending to one's experience (Burge, 2006). One can have a conscious experience of red, and that experience can have whatever awareness comes with conscious experience, even in the absence of top-down attention to it (Koch & Tsuchiya, 2007). Another rival to the higher order account of why a conscious state is one that one is conscious of is the *same order* account in which

a conscious pain is reflexive in that it is about itself. That is, it has a content that turns back on itself, and that is what makes a pain a state one is conscious of. This view had its beginnings in Aristotle (Caston, 2002) and was later pursued by Brentano (1874/1973). (See Burge, 2006; Kriegel & Williford, 2006.) Either one of the deflationary or same order accounts can be adopted by advocates of the biological view and the global workspace view, so I see no real advantage for the HOT view here.

Further problems for the HOT theory

I argued that the HOT theory cannot recognize an explanatory gap, but my argument was oversimple because it neglected a crucial distinction between two types of HOT theories. The kind of HOT theory that cannot recognize an explanatory gap is the *ambitious* HOT theory of *phenomenal* consciousness that analyzes *phenomenal* consciousness in terms of higher order thought. But there is also a *modest* and therefore innocuous form of the HOT theory that just says that, in addition to phenomenal consciousness, there is *another kind of consciousness*, higher order consciousness. Phenomenal consciousness is one thing, and higher order consciousness is another. The modest form can recognize an explanatory gap for phenomenal consciousness. The modest account is suggested by Lycan's remark, "I cannot myself hear a natural sense of the phrase 'conscious state' other than as meaning 'state one is conscious of being in'" (Lycan, 1996b). As Lycan recognizes, what one can and cannot "hear" leaves the theoretical options open. The modest account is tantamount to a verbal claim—that there is a sense of the term "conscious" (distinct from "phenomenal consciousness") that has a higher order meaning—and does not dictate that there is no explanatory gap. The very short argument against the HOT theory (that it does not recognize an explanatory gap and so is false) is an argument only against the *ambitious* form of the HOT theory. In the rest of this section, I will explain some other problems with the ambitious HOT theory that also do not apply to the *modest* version.

The first thing to realize about the HOT theory in both the ambitious and modest forms is that it needs considerable qualification. Suppose I consciously infer that I am angry from my angry behavior, or—in a slightly different kind of case that need not involve conscious inference—I am aware of my anger in noticing my angry fantasies. In these cases we would not say the anger is thereby conscious. Further, Freudians sometimes suppose that a subject can unconsciously recognize his own desire to, for example, kill his father and marry his mother, along with the need to cloak that desire in a form that will not cause damage to the self. But we would not say that in virtue of such an unconscious HOT (one that cannot readily become conscious) about it, the desire is therefore conscious! These examples concerning what we would say suggest that a HOT about a state is not something we regard as sufficient for the state to be conscious. Defenders of the HOT theory introduce *complications* in the HOT theory to try to avoid these counterexamples. Rosenthal (2005a) says that S is a conscious state if and only if S is accompanied by a thought to the effect that the subject is in S *that is arrived at without inference or observation of which the subject is conscious*. The italicized phrase avoids the problems posed by conscious observation of angry fantasies and conscious inference by stipulating that HOTs arrived at by conscious observation and inference are not sufficient for consciousness. (Another stipulation that I will not describe is supposed to handle the Freudian issue.) Suppose as a result of biofeedback training I come to have noninferential nonobservational knowledge of states of my liver (Block, 1995). Since we would not count the state of the liver as conscious in virtue of the HOT about it, Rosenthal (2000b, p. 240) further stipulates that only mental states can be conscious. What if I have a HOT about my future or past mental state? Rosenthal (2000b, p. 241) further stipulates that if one has a thought about a state, that makes it conscious only when one thinks of it as present to oneself.

As Bertrand Russell noted in an often-quoted passage (1919, p. 71), "The method of 'postulating' what we want has many advantages; they are the same as the advantages of theft over honest toil." Honest toil is not required if the HOT view is understood as a modest account, since *stipulation is not a problem in a stipulated sense of a term*, but ad hoc stipulation is a problem if we take the HOT view as an ambitious account, especially as an empirical theory of consciousness.

A second class of issues concerns the "mismatch problem," the possibility of a mismatch in content between a sensory representation and the accompanying HOT. What phenomenally conscious quality does an experience have if a HOT to the effect that one has a dull throbbing pain in the toe is accompanied not by any representation of toe damage but instead a visual representation of red—or by no sensory representation at all? If the sensory representation determines the conscious quality all by itself, the contents of HOTs are irrelevant here, and if here, why not elsewhere? And if the HOT determines the conscious quality without the sensory representation, then the contents of sensory representations are irrelevant—so what is the difference between *thinking* you have a delightful experience and actually having one (Byrne, 1997; Neander, 1998; Balog, 2000; Rey, 2000; Levine, 2001)? Of course, new sophistication in one's HOTs, as when one learns to recognize different wines, can cause a corresponding differentiation in the sensory states that the HOTs are about, but HOTs are not always causally self-fulfilling (if only!), and in any case, causal self-fulfillment does not answer the *constitutive* question of what the difference is between thinking you have an experi-

ence of a certain sort and actually having one. Rosenthal (2000b; 2000a; 2005b, pp. 217–219) claims that a HOT is sufficient for a conscious state even without any sensory representation that the HOT is about. But suppose I have a sharp pain that causes a HOT to the effect that I have a sharp pain, through the normal processes by which pains often cause metacognitions about them. And suppose that by chance I also have a qualitatively different sharp pain (one pain is a bit sharper than the other) that produces no HOT at all. The content of the HOT—that I have a sharp pain—does not distinguish between the two pains even though by any ordinary standard it is about one of them but not the other. If the HOT theory follows common sense, saying that one pain is conscious but the other is not, it is hard to see how that (partly causal) way of cashing out aboutness could be compatible with the claim that a HOT to the effect that I am in pain could be a conscious pain on its own without any sensory representation.

A third class of issues concerns children. If you have seen and heard a circumcision, you may find it difficult to doubt that it hurts. Relevant evidence: newborns who are circumcised without anesthesia or analgesia are more stressed by later vaccination even 6 months later (Taddio, Goldbach, Ipp, Stevens, & Koren, 1995). My point is not that you should be totally convinced of phenomenal consciousness in early infancy, but rather that you should be convinced that there is a *better* case for phenomenal consciousness in infancy than there is for those instances of phenomenal consciousness being accompanied by higher order thought.

One point against higher order thought in infancy is that frontal cortex, the likely neural home of thought about thought (Stone, Baron-Cohen, & Knight, 1998) is immature in infancy. Gazzaniga, Ivry, and Mangun (2002, pp. 642–643) discuss two sources of evidence that areas of the brain that specialize in sensory and motor function develop significantly earlier than areas responsible for thinking. One source of evidence derives from autopsy results on human brains from age 28 weeks after conception to 59 years of age. The result, diagrammed in figure 77.2, is that auditory synaptic density peaks at about 3 months (and probably likewise for synaptic density in other sensory areas), whereas the association areas of the frontal cortex peak at about 15 months. Similar results apply to PET imaging, which measures glucose metabolism in different parts of the brain.

As infants become more mature, our confidence in their phenomenal consciousness increases, as does our confidence in their capacity for higher order thought. However, it continues to be doubtful that phenomenally conscious states are always accompanied by higher order thoughts. Children even up to age 3–4 have difficulty thinking about their own states of mind. For example, Alison Gopnik and her colleagues (Gopnik & Graf, 1988) used a tube that was open at both ends and contained a window that could be open or closed. The child would be asked to either look in the window or reach into the side and identify a common object, for example, a spoon. Then with the apparatus taken away, the

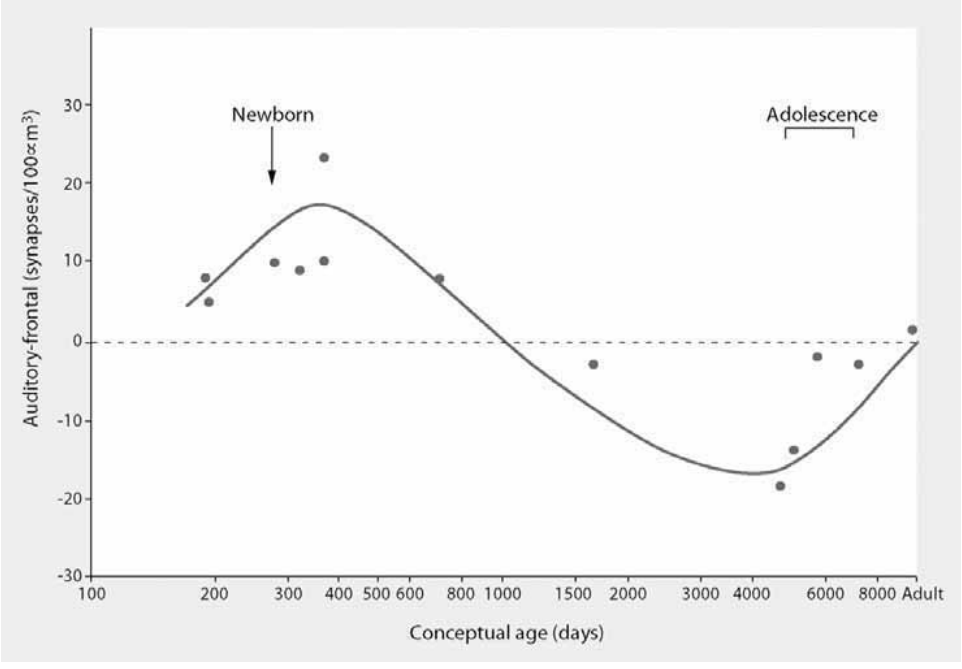


Figure 77.2 Relative synaptic density of auditory and frontal cortex. Conceptual age is age from conception. The peak at the left of roughly three months (postnatal) reflects a high number of auditory synapses relative to frontal synapses. (From Gazzaniga, Ivry, & Mangun, 2002.)

child was asked how he or she knew the spoon was in the tube. The children were nearly random in their answers, probably because, as Gopnik has pointed out in a series of papers (see Gopnik, 2007), they have difficulty attending to and thinking about their own representational states. Marjorie Taylor and her colleagues have compared "source amnesia" for representational states of mind with skills (Esbensen, Taylor, & Stoess, 1997). For example, some children were taught to count in Japanese, whereas other children were taught the Japanese word for "three." Children were much less likely to be able to name the source of their representational state than the source of their counting skill. (For example, "You just taught me" in answer to the skill question versus "I've always known" in answer to the representational state question.) The source amnesia results apply most directly to conscious intentional states rather than conscious perceptual states, but to the extent that perceptual states are representational, they may apply to them as well. Older autistic children who clearly have phenomenally conscious states also have problems attending to and thinking about representational states of mind (Baron-Cohen, 1995; Charman & Baron-Cohen, 1995). Will a defender of the ambitious HOT theory tell us that these autistic children lack phenomenal states? Or that contrary to the evidence they do have HOT states?

I emphasize that it is difficult for young children and autists to think about representational states of mind—but not impossible. Indeed, children as young as 13 months can exhibit some ability to track others' beliefs (Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007). In the case of false belief, as in many other examples of cognition, a cognitive achievement is preceded by a highly modular and contextualized analog of it, one that partly explains the development of the cognitive achievement. My point is not that metacognition in all its forms is impossible in young children and autists but that, at all ages, our justification for attributing conscious states exceeds our justification for attributing metacognitive states.

Although the empirical case against the higher-order-thought point of view is far from overwhelming, it is strong enough to make the question salient of what the advantages of the *ambitious* higher-order-thought theory of consciousness actually are (as contrasted with the advantages of the modest version, which none of these points apply to).

But how do we know whether a version of the HOT theory is ambitious or modest? One way to tell is to ask whether, on that theory, a phenomenally conscious state—considered independently of any HOT about it—is something that is bad or good in itself. For example, Carruthers (1989, 1992) famously claimed that because pains in dogs, cats, sheep, cattle, pigs, and chickens are not available to be thought about, they are not felt and hence not anything to be concerned about; that is, they are states with no moral significance. (Carruthers later, in 1999, took a different view on the grounds that frustration of animal desires is of moral significance even though the pains themselves are not.)

I turn now to related issues about the self that may seem to go against the biological view.

The self

The biological view may seem at a disadvantage with respect to the self. Since Hume (1740/2003) described the self as "that connected succession of perceptions," many (Dennett, 1984; Parfit, 1984) have thought about persons in terms of integrated relations among mental states. The global workspace view seems well equipped to locate consciousness as self-related given that broadcasting in the global workspace is itself a kind of integration. And the HOT view at least requires the integration of one state being about another. By contrast, it looks as if, on many views of the biological basis of a conscious state (Block, 1995), it could exist without integration, and this point has resulted in accusations of scanting the self (Church, 1995; Harman, 1995; Kitcher, 1995). One response would be to favor a biological neural basis of consciousness that itself involves integration (Tononi & Edelman, 1998; Tononi & Koch, 2008). But it is worth pointing out that phenomenal consciousness has less to do with the self than critics often suppose.

What is the relation between phenomenal consciousness and the self? We could raise the issue by thinking about pain asymbolia, a syndrome in which patients have pain experiences without the usual negative affect (Aydede, 2005): they do not seem to mind the pain. In this syndrome, patients sometimes describe the pains as painful for someone else, and perhaps they are right given pain's unusual lack of connection to the subject's emotions, planning, and valuation. Here is a question about such a dissociation syndrome: If such a subject thinks about the painfulness of such a pain (as opposed to its merely sensory aspect), is the painfulness thereby phenomenally conscious? It would seem not, suggesting that the kind of integration supplied by HOTs is not actually sufficient for consciousness.

Here is another conundrum involving the relation between phenomenal consciousness and the self. In many experiments, activation in the fusiform face area at the bottom of the temporal lobe has been shown to correlate with the experience of a face. Now, injury to the parietal lobe often causes a syndrome called visuospatial extinction. If the patient sees a single object, the patient can identify it, but if there are objects on both the right and the left, the patient claims not to see one—most commonly the one on the left. However two fMRI studies (Rees et al., 2000; Rees, Wojciulik, et al., 2002) have shown that in patient GK, when GK claims not to see a face on the left, his fusiform face area lights up almost as much as when he reports seeing the

face. One possibility is that the original identification of the fusiform face area as the neural basis of face experience was mistaken. But another possibility is that the subject genuinely has face experience that he does not know about and cannot know about. Wait—is that really a possibility? Does it even make sense to suppose that a subject could have an experience that he does not and cannot know about? What would make it *his* experience?

The question about GK can be answered by thinking about the subject's visual field. We can answer the question of what the visual field is by thinking about how it is measured. If you look straight ahead, hold a rod out to the side, and slowly move it forward, you will be able to see it at roughly 100° from the forward angle. If you do the same coming down from the top, you will see it at roughly 60°, and if you do it from the bottom, you will see it at roughly 75°. More accurately, it is measured with points of light or gratings. Thus the visual field is an oval, elongated to the right and left, and slightly larger on the bottom. The Humphrey Field Analyzer HFA-II-I can measure your visual field in as little as 2 minutes. The United Kingdom has a minimum visual field requirement for driving (60° to the side, 20° above and below); U.S. states vary widely in their requirements (Peli & Peli, 2002). I mention these details to avoid skepticism about whether the visual field is real.

The visual field can help us think about GK. If GK does genuinely experience the face on the left that he cannot report, then it is in his visual field on the left side, and as such has relations to other items in his visual field, some of which he will be able to report. The fact that it is his visual field shows that it is *his* experience. I caution the reader that this discussion concerns the issue of whether it *makes sense* to describe GK as having an experience that he cannot know about and does not constitute any evidence for his actually having an experience that he cannot know about (but see Block, 2007a).

A second point about the relation between phenomenal consciousness and the self is that self-related mental activities seem inhibited during intense conscious perception. Malach and his colleagues (Goldberg, Harel, & Malach, 2006) showed subjects pictures and audio clips with two different types of instructions. In one version, subjects were asked to indicate their emotional reactions as positive, negative, or neutral. In another version (in which the stimuli were presented much faster), subjects were asked to categorize the stimuli, for example, as animals or not. Not surprisingly, subjects rated their self-awareness as high in the introspective task and low in the categorization task. And this testimony was supported by fMRI results that showed that the introspective task activated an "intrinsic system" that is linked to judgments about oneself, whereas the categorization task inhibited the intrinsic system, activating instead an extrinsic system that is also activated when subjects viewed clips from Clint Eastwood's *The Good, the Bad, and the Ugly*. Of course this result does not show that intense perceptual experiences are not part of a connected series of mental states constituting a self, but it does suggest that theories that bring any *sense of self* into phenomenal experience are wrongheaded. Malach's result disconfirms the claim that a conscious visual experience consists in a perceptual state causing a thought to the effect that *I myself* have a visual experience (Rosenthal, 2005a).

Machine consciousness

The global workspace account lends itself particularly well to the idea of machine consciousness. There is nothing intrinsically biological about a global workspace. And the HOT view also is friendly to machine consciousness. If a machine can think, and if it can have representational contents, and if it can think about those contents, it can have conscious states, according to the HOT view. Of course, we do not know how to make a machine that can think, but whatever difficulties are involved in making a machine think, they are not difficulties about consciousness per se. (However, see Searle, 1992, for a contrary view.) By comparison, the biological theory says that only machines that have the right biology can have consciousness, and in that sense the biological account is less friendly to machine consciousness. Information is coded in neurons by electrical activations that travel from one part of a neuron to another, but in the most common type of transfer of information between neurons, that electrical coding is transformed into a chemical coding (by means of neurotransmitters) which transfers the information to another neuron where the coding of information is again electrical. On the biological view, it may well be that this transfer of coding of information from electrical to chemical and back to electrical is necessary to consciousness. Certainly it would be foolish to discount this possibility without evidence.

As should be apparent, the competitors to the biological account are profoundly nonbiological, having more of their inspiration in the computer model of the mind of the 1960s and 1970s than in the age of the neuroscience of consciousness of the 21st century. (For an example, see McDermott, 2001.) As Dennett (2001, 234) confesses, "The recent history of neuroscience can be seen as a series of triumphs for the lovers of detail. Yes, the specific geometry of the connectivity matters; yes, the location of specific neuromodulators and their effects matter; yes, the architecture matters; yes, the fine temporal rhythms of the spiking patterns matter, and so on. Many of the fond hopes of opportunistic minimalists [a version of computationalism: NB] have been dashed: they had hoped they could leave out various things, and they have learned that no, if you leave out *x*, or *y*, or *z*, you can't explain how the mind works." Although Dennett resists the

obvious conclusion, it is hard to avoid the impression that the biology of the brain is what matters to consciousness—at least the kind we have—and that observation favors the biological account.

ACKNOWLEDGMENTS I am grateful to Susan Carey, Peter Carruthers, Christof Koch, David Rosenthal, and Stephen White for comments on an earlier version.

REFERENCES

ALKIRE, M. T. (2008). General anesthesia and consciousness. In S. Laureys (Ed.), *The neurology of consciousness: Cognitive neuroscience and neuropathology*: Amsterdam: Elsevier.

ALKIRE, M. T., HAIER, R. J., & FALLON, J. H. (2000). Toward a unified theory of narcosis: Brain imaging evidence for a thalamocortical switch as the neurophysiologica basis of an anesthetic-induced unconsciousness. *Conscious. Cogn.*, *9*(3), 370–386.

ALKIRE, M. T., & MILLER, J. (2005). General anesthesia and the neural correlates of consciousness. *Prog. Brain Res.*, *150*, 229–244.

ARMSTRONG, D. M. (1978). What is consciousness? *Proc. Russellian Soc.*, *3*, 65–76.

AYDEDE, M. (2005). A critical and quasi-historical essay on theories of pain. In M. Aydede (Ed.), *Pain: New essays on its nature and the methodology of its study*. Cambridge, MA: MIT Press.

BAARS, B. (1988). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.

BALOG, K. (2000). Phenomenal judgment and the HOT theory: Comments on David Rosenthal's "Consciousness, content and metacognitive judgments." *Conscious. Cogn.*, *9*(2), 215–218.

BARON-COHEN, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.

BLOCK, N. (1978). Troubles with functionalism. *Minn. Stud. Philos. Sci.*, *9*, 261–325.

BLOCK, N. (1980). What is functionalism? In N. Block (Ed.), *Readings in the philosophy of psychology* (pp. 171–184). Cambridge MA: Harvard University Press.

BLOCK, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.*, *18*(2), 227–247.

BLOCK, N. (2006). Max Black's objection to mind-body identity. *Oxf. Stud. Metaphys.*, *2*, 3–78.

BLOCK, N. (2007a). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.*, *30*, 481–548.

BLOCK, N. (2007b). Overflow, access and attention. *Behav. Brain Sci.*, *30*, 530–542.

BRENTANO, F. (1874/1973). *Psychology from an empirical standpoint* (A. Rancurello, D. B. Terrell, & L. L. McAlister, Trans.). London: Routledge and Kegan Paul.

BRITTEN, K. H., SHADLEN, M. N., NEWSOME, W. T., & MOVSHON, A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J. Neurosci.*, *12*, 4745–4765.

BURGE, T. (2006). Reflections on two kinds of consciousness. In *Philosophical essays*, Vol. 2: *Foundations of mind* (pp. 392–419). New York: Oxford University Press.

BYRNE, A. (1997). Some like it HOT: Consciousness and higher-order thoughts. *Philos. Stud.*, *86*, 103–129.

BYRNE, A. (2001a). Intentionalism defended. *Philos. Rev.*, *110*, 199–240.

BYRNE, A. (2001b). Review of Carruthers' phenomenal consciousness. *Mind*, *110*, 1057–1062.

CARRUTHERS, P. (1989). Brute experience. *J. Philos.*, *86*, 258–269.

CARRUTHERS, P. (1992). *The animals issue: Moral theory in practice*. Cambridge, UK: Cambridge University Press.

CARRUTHERS, P. (1999). Sympathy and subjectivity. *Australas. J. Philos.*, *77*, 465–482.

CARRUTHERS, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge, UK: Cambridge University Press.

CASTON, V. (2002). Aristotle on consciousness. *Mind*, *111*(444), 751–815.

CHALMERS, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford, UK: Oxford University Press.

CHARMAN, T., & BARON-COHEN, S. (1995). Understanding models, photos and beliefs: A test of the modularity thesis of metarepresentation. *Cogn. Dev.*, *10*, 287–298.

CHURCH, J. (1995). Fallacies or analyses? *Behav. Brain Sci.*, *18*(2), 251–252.

CHURCHLAND, P. (2002). *Brain-wise: Studies in neurophilosophy*. Cambridge MA: MIT Press.

COWEY, A., & WALSH, V. (2000). Magnetically induced phosphenes in sighted, blind and blindsighted subjects. *NeuroReport*, *11*, 3269–3273.

CRANE, T. (2000). The origins of qualia. In T. Crane & S. Patterson (Eds.), *History of the mind-body problem* (pp. 169–194). New York: Routledge.

CRICK, F. (1994). *The astonishing hypothesis*. New York: Scribners.

CRICK, F., & KOCH, C. (1990). Towards a neurobiological theory of consciousness. *Sem. Neurosci.*, *2*, 263–275.

CRICK, F., & KOCH, C. (1998). Consciousness and neuroscience. *Cereb. Cortex*, *8*, 97–107.

CURTIS, C., & D'ESPOSITO, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.*, *7*(9), 415–423.

DAMASIO, A., & MEYER, K. (2008). Behind the looking-glass. *Nature*, *454*, 167–168.

DEHAENE, S., CHANGEUX, J.-P., NACCHACHE, L., SACKUR, J., & SERGENT, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends Cogn. Sci.*, *10*, 204–211.

DENNETT, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: MIT Press.

DENNETT, D. (2001). Are we explaining consciousness yet. *Cognition*, *79*, 221–237.

DRETSKE, F. (1993). Conscious experience. *Mind*, *102*, 263–283.

ESBENSEN, B. M., TAYLOR, M., & STOESS, C. J. (1997). Children's behavioral understanding of knowledge acquisition. *Cogn. Dev.*, *12*, 53–84.

FEIGL, H. (1958). The "mental" and the "physical." *Minn. Stud. Philos. Sci.*, *2*, 370–497.

GALLISTEL, C. R. (1998). Insect navigation: Brains as symbol-processing organs. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues;* Vol. 4 of *An invitation to cognitive science* (2nd ed.). Cambridge, MA: MIT Press.

GAZZANIGA, M., IVRY, R., & MANGUN, G. (2002). *Cognitive neuroscience: The biology of the mind*. New York: W. W. Norton.

GOLDBERG, I. I., HAREL, M., & MALACH, R. (2006). When the brain loses its self: Prefrontal inactivation during sensorimotor processing. *Neuron*, *50*, 329–339.

GOPNIK, A. (2007). Why babies are more conscious than we are. *Behav. Brain Sci.*, *30*(5), 503–504.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Children's understanding of the sources of their knowledge. *Child Dev.*, *59*, 1366–1371.

Harman, G. (1995). Phenomenal fallacies and conflations. *Behav. Brain Sci.*, *18*(2), 256–257.

Heeger, D. J., Boynton, G. M., Demb, J. B., Seideman, E., & Newsome, W. T. (1999). Motion opponency in visual cortex. *J. Neurosci.*, *19*, 7162–7174.

Hobbes, T. (1989). *Metaphysical writings of Thomas Hobbes*. LaSalle, IL: Open Court.

Horgan, T., & Tienson, J. (2001). Deconstructing new wave materialism. In C. Gillett & B. Loewer (Eds.), *Physicalism and its discontents*. New York: Cambridge University Press.

Huk, A. C., Ress, D., & Heeger, D. J. (2001). Neuronal basis of the motion aftereffect reconsidered. *Neuron*, *32*, 161–172.

Hume, D. (1740/2003). *A treatise of human nature*. New York: Dover.

Jackson, F. (1982). Epiphenomenal qualia. *Am. Philos. Q.*, *32*, 127–136.

Kammer, T. (1999). Phosphenes and transient scotomas induced by magnetic stimulation of the occipital lobe: Their topographic relationship. *Neuropsychologia*, *37*, 191–198.

Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition*, *79*, 89–113.

Kirk, G. S., Raven, J. E., & Schofield, M. (1983). *The presocratic philosophers* (2nd ed.). Cambridge, UK: Cambridge University Press.

Kitcher, P. (1995). Triangulating phenomenal consciousness. *Behav. Brain Sci.*, *18*(2), 266-167.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Englewood, CO: Roberts.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends Cogn. Sci.*, *11*, 16–22.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *J. Cogn. Neurosci.*, *12*, 48–55.

Kriegel, U., & Williford, K. (2006). *Self-representational approaches to consciousness*. Cambridge, MA: MIT Press.

Lamme, V. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci.*, *7*, 12–18.

Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends Cogn. Sci.*, *9*(2), 556–559.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pac. Philos. Q.*, *64*, 354–361.

Levine, J. (2001). *Purple haze: The puzzle of consciousness*. Oxford, UK: Oxford University Press.

Loar, B. (1990/1997). Phenomenal states. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 597–616). Cambridge MA: MIT Press.

Lycan, W. G. (1996a). *Consciousness and experience*. Cambridge, MA: MIT Press.

Lycan, W. G. (1996b). Consciousness as internal monitoring. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical and scientific debates* (pp. 755–772). Cambridge MA: MIT Press.

Macknik, S. L., & Martinez-Conde, S. (2007). The role of feedback in visual masking and visual processing. *Advances in Cognitive Psychology*, *3*(1–2), 125–152.

McDermott, D. (2001). *Mind and mechanism*. Cambridge, MA: MIT Press.

McGinn, C. (1991). *The problem of consciousness*. Oxford, UK: Oxford University Press.

Naccache, L., & Dehaene, S. (2007). Reportability and illusions of phenomenality in the light of the global neuronal workspace model. *Behav. Brain Sci.*, *30*, 518–519.

Nagel, T. (1974). What is it like to be a bat? *Philos. Rev.*, *83*(4), 435–450.

Neander, K. (1998). The division of phenomenal labor: A problem for representational theories of consciousness. *Philos. Perspect.*, *12*, 411–434.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258.

Papineau, D. (2002). *Thinking about consciousness*. New York: Oxford University Press.

Parfit, D. (1984). *Reasons and persons*. Oxford, UK: Oxford University Press.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*, 510–512.

Peli, E., & Peli, D. (2002). *Driving with confidence: A practical guide to driving with low vision*. Singapore: World Scientific.

Place, U. T. (1956). Is consciousness a brain process? *Br. J. Psychol.*, *47*, 44–50.

Prinz, J. (2007). Accessed, accessible, and inaccessible: Where to draw the phenomenal line. *Behav. Brain Sci.*, *30*, 521–522.

Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.*, *3*, 261–270.

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C., & Driver, J. (2000). Unconscious activations of visual cortex in the damaged right-hemisphere of a parietal patient with extinction. *Brain*, *123*(8), 1624–1633.

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C., & Driver, J. (2002). Neural correlates of conscious and unconscious vision in parietal extinction. *Neurocase*, *8*, 387–393.

Rey, G. (2000). Role, not content: Comments on David Rosenthal's "Consciousness, content and metacognitive judgments." *Conscious. Cogn.*, *9*(2), 224–230.

Rosenberg, G. (2004). *A place for consciousness: Probing the deep structure of the natural world*. New York: Oxford University Press.

Rosenthal, D. (2000a). Consciousness, content and metacognitive judgments. *Conscious. Cogn.*, *9*(2), 203–214.

Rosenthal, D. (2000b). Metacognition and higher-order thoughts. *Conscious. Cogn.*, *9*(2), 231–242.

Rosenthal, D. (2005a). *Consciousness and mind*. New York: Oxford University Press.

Rosenthal, D. (2005b). Sensory qualities, consciousness and perception. In *Consciousness and mind* (pp. 175–226). Oxford, UK: Oxford University Press.

Russell, B. (1919). *Introduction to mathematical philosophy*. New York: Routledge.

Schneider, G., & Kochs, E. F. (2007). The search for structures and mechanisms controlling anesthesia-induced unconsciousness. *Anesthesiology*, *107*(2), 195–198.

Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Sergent, C., & Rees, G. (2007). Conscious access overflows overt report. *Behav. Brain Sci.*, *30*, 523–524.

Siegel, S. (2008). The contents of perception. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. http://plato.stanford.edu/archives/fall2008/entries/perception-contents/

Silvanto, J., Cowey, A., Lavie, N., & Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. *Nat. Neurosci.*, *8*(2), 143–144.

Smart, J. J. C. (1959). Sensations and brain processes. *Philos. Rev.*, *68*, 141–156.

Sosa, E. (2003). Privileged access. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives*. Oxford, UK: Oxford University Press.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci.*, *10*(5), 640–656.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychol. Sci.*, *18*(7), 580–586.

Taddio, A., Goldbach, M., Ipp, M., Stevens, B., & Koren, G. (1995). Effect of neonatal circumcision on pain responses during vaccination in boys. *Lancet*, *345*(8945), 291–292.

Théoret, H., Kobayashi, M., Ganis, G., Di Capua, P., & Pascual-Leone, A. (2002). Repetitive transcranial magnetic stimulation of human area MT/V5 disrupts perception and storage of the motion aftereffect. *Neuropsychologia*, *40*(13), 2280–2287.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, *282*, 1846–1851.

Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: An update. *Ann. NY Acad. Sci.*, *1124*, 239–261.

Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.

Velly, L., Rey, M., Bruder, N., Gouvitsos, F., Witjas, T., Regis, J., et al. (2007). Differential dynamic of action on cortical and subcortical structures of anesthetic agents during induction of anesthesia. *Anesthesiology*, *107*(2), 202–212.

White, S. (2006). A posteriori identities and the requirements of rationality. *Oxf. Stud. Metaphys.*, *2*, 91–102.

Zihl, J., von Cramon, D., & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain*, *106*, 313–340.

78 Recovery of Consciousness after Brain Injury: An Integrative Research Paradigm for the Cognitive Neuroscience of Consciousness

NICHOLAS D. SCHIFF

ABSTRACT This chapter reviews the evolving understanding of recovery of consciousness following severe brain injuries. Important questions guiding current research are considered including the need to develop new diagnostic tools based on neuroimaging methods that can guide longitudinal assessments of brain function. Novel assessments of cognitive function in the absence of behavior and slowly evolving plastic changes in brain structure that may arise after injury are reviewed. Emphasis is placed on the importance of developing more precise and testable models at the "circuit level" that are predictive of patterns of recovery and response to treatments. It is argued that recovery of consciousness in the human brain is likely to depend on interactions between dynamic circuit-level mechanisms that support reestablishment of goal-directed behaviors and cellular repair mechanisms. Further suggestions for the development of a cognitive neuroscience of the recovery of consciousness based on quantitative assessment of evolving changes in behavior and brain function are outlined.

Despite major advances in neuroscience, recovery of consciousness after brain injuries remains poorly understood. At the origin of this challenge is the surprisingly wide range of underlying brain function that may be present when confronted by a patient at the bedside with very limited or even no overt signs of behavioral responsiveness. Across the range of behavioral features consistent with clinical diagnoses ranging from vegetative state (no evidence of self or environmental awareness), minimally conscious state (at least some evidence of awareness), and up to but not including patients in locked-in state (full consciousness with no motor control), there are many patients whose level of consciousness we cannot at present confidently assess. This chapter reviews several recent studies that have expanded our understanding of this problem and the fascinating conceptual challenges it presents for a cognitive neuroscience of consciousness. As will be discussed in greater detail, some patients with clinical examination features consistent with vegetative state, or demonstrating only the lowest level of nonreflexive motor responses indicating minimally conscious state (MCS), can be shown through the use of neuroimaging techniques to be able to follow commands and possibly have higher cognitive capacities. Other patients who show no functional recovery beyond MCS for five, ten, or twenty years may have substantial spontaneous (or induced) recoveries of spoken language and cognitive functions. Such patients have remained amnestic for the decades of intervening time while charting new and stable patterns of recovery. Other patients may recover memory and cognition but remain unable to control intact motor pathways, so that they cannot communicate their awareness to the people in their environment.

Among the central concerns of current research aimed at understanding the recovery of consciousness after brain injury are the following questions: Can functional neuroimaging techniques reliably, and operationally, identify levels of awareness, memory, and other higher brain functions in patients who do not show behavioral evidence of these cognitive capacities? What role do changes in brain structure and brain dynamics play in the recovery process? How do such changes, if present, evolve over time and develop? Can the natural recovery process be impeded by abnormal functional activity or be facilitated by interventions? As further developed in this chapter, several recent observations reveal hints of common "circuit-level" mechanisms linked to the impairment and recovery of human consciousness. Specific observations of the process of recovery of consciousness after

NICHOLAS D. SCHIFF Laboratory for Cognitive Neuromodulation, Department of Neurology and Neuroscience, Weill Medical College, Cornell University, New York, New York

brain injuries also point to a related set of biological mechanisms associated with changes in brain structure and intrinsic cellular function.

Disorders of consciousness: An overview

Figure 78.1 diagrams the varying relationships between impaired cognitive function and motor function in human disorders of consciousness. From the bottom left of figure 78.1, the conditions of coma and vegetative state (VS) are both considered unconscious brain states as judged behaviorally by complete unresponsiveness to environmental stimuli and lack of self-initiated behavior despite integrity of the motor pathways that could enable such action. Importantly, these clinical conditions are not diagnosed solely on behavioral criteria, and the natural history of the type of brain injury producing the condition is generally more important to the diagnostic process (see Posner, Saper, Schiff, & Plum, 2007; Schiff, 2007). Whereas comatose patients show no variations in state, and typically their eyes remain closed and they show no response to the most vigorous stimulation, VS patients recover a crude sleep-wake cycling reflected in irregular periods of eye opening and eye closure. This cyclical variation in VS does not correlate with identifiable electroencephalographic (EEG) features of either sleep or normal wakefulness; permanent vegetative state is typically associated with low-frequency, monotonous EEG signals, whereas MCS patients generally show preservation of many key features of the normal structure of the sleep-wake EEG (Kobylarz & Schiff, 2005).

To the right of VS on figure 78.1 is a gray zone indicating a transitional group of patients with atypical clinical features who nonetheless show no response to external stimulation or evidence of intentional behaviors (see Schiff, 2004; Schiff, Ribary, et al., 2002). Once unequivocal but inconsistent evidence of awareness of self or the environment can be demonstrated at the bedside, patients enter into the minimally conscious state (MCS). MCS patients may show a wide range of clinical features (Giacino & Whyte, 2005) with the upper boundary determining a patient's emergence from MCS based on the recovery of reliable verbal or gestural communication. At present, there is no predictive time frame for emergence from MCS following severe brain injuries with rare examples of full recovery of fluent spoken language after more than ten years (see the section "Association between functional recovery and white matter structural changes"). Studies of functional outcome five years after injury demonstrate a lack of correlation of time in MCS and ultimate level of recovery (Lammi, Smith, Tate, & Taylor, 2005).

As seen in figure 78.1, it is possible that a patient with no controllable motor output channel may be fully conscious yet display a behavioral profile consistent with deep coma: eyes closed and unresponsive to any external stimuli as judged by a bedside examination. The locked-in state (LIS) at the far right bottom of the figure defines this condition and is not a disorder of consciousness. LIS patients

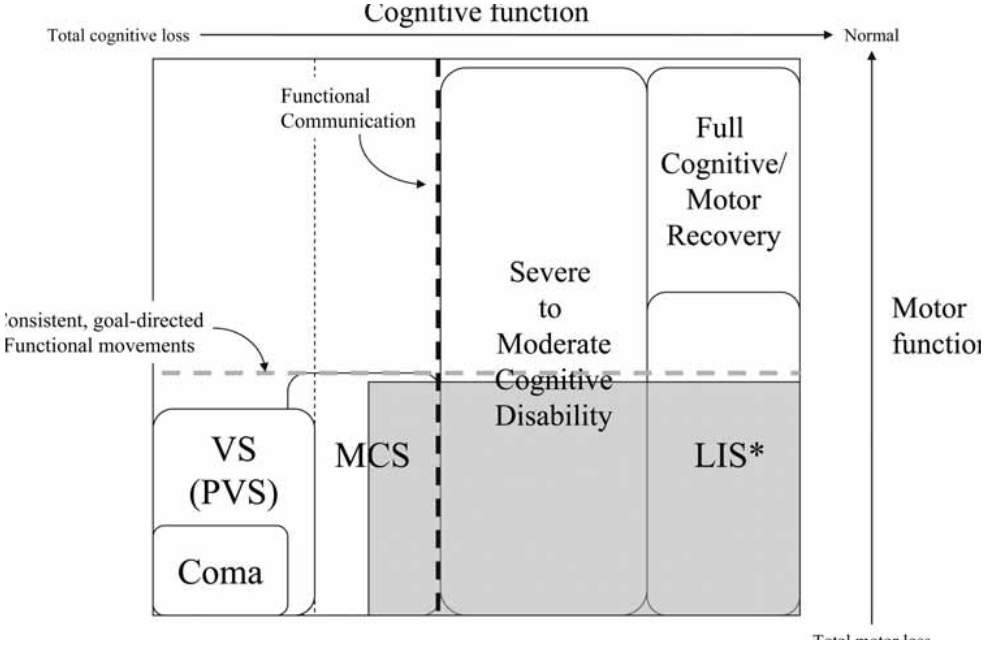


FIGURE 78.1 Correspondence of cognitive and motor impairment across human disorders of consciousness. VS, vegetative state; MCS, minimally conscious state; LIS, locked-in state, which is not a disorder of consciousness.

retain total preservation of cognitive function but may have no or little motor function. Those that have no motor function whatsoever are represented below the horizontal interrupted gray line in figure 78.1. Many LIS patients, however, can signal using eye movements and sometimes lateral head movements, and this group of LIS patients is represented above this line (Laureys et al., 2005). Brain injuries producing LIS typically involve the ventral pontine regions. Because of initial swelling causing transient dysfunction of nearby brain stem neurons that maintain arousal level (located in the dorsal pons and midbrain), LIS patients typically experience an initial comatose phase at onset of injury.

While a natural history of a neurological disorder that selectively disrupts the motor pathways or slowly erodes motor function may lead to the appropriate prospective expectation that the subject is conscious, the complexity of many brain injuries can make this determination highly uncertain. The large gray box covering the upper range of MCS and extending to the extreme left of the diagram identifies the most problematic set of patients where motor function is so severely impaired as to prohibit consistent goal-directed movements that allow for communication. Patients who retain significant cognitive capacity or remain within the normal range of cognitive function may not be recognizably different from MCS patients if motor impairment places them within this region of the figure. The uncertainty present in the assessment of such patients is perhaps the most difficult area for accurate measurement of brain state and cognitive capacity and indicates where evolving use tools of cognitive and clinical neuroscience will have their greatest impact. The variations of brain function and behavior seen between patients in vegetative state and minimally conscious state are treated elsewhere (see Schiff, 2004, 2005, 2006).

Between MCS and LIS, patients may retain varying levels of cognitive processing capabilities, awareness, memory, and other higher brain functions. If motor function is sufficiently impaired such that patients cannot reliably signal through controlled goal-directed movements (dashed horizontal line), it is not possible even using quantitative behavioral assessments to judge their cognitive capacity or communicate with them. Thus disentangling a patient's potential for cognitive function from limitations due to dysfunction of internal motor control systems and sensorimotor integration mechanisms presents the main challenge to understanding the nature of patients' interior states and developing strategies to help them. This diagnostic challenge is now being addressed by development of hierarchical neuroimaging protocols to systematically assess the integrity of sensory processing systems and assess the capacity to follow commands or establish a communication channel in the absence of overt behavior (Owen et al., 2006; Coleman et al., 2007; Kübler & Kotchoubey, 2007).

Detecting awareness in the absence of behavioral responsiveness

Using a novel functional magnetic resonance imaging (fMRI) assessment paradigm, Owen and colleagues provided the first evidence that neuroimaging can be used in the absence of a visible behavioral response to obtain unambiguous evidence of the conscious responses indicated by command following (Owen et al., 2006). Typically, simple command following is assessed at a patient's bedside by observation of small movements such as finger, thumb, or eye movements in response to an examiner's direction. Reliable command following without the concomitant ability to use such small movements to establish a communication system is seen in many MCS patients (Giacino & Whyte, 2005). In a study of a patient who had remained in a diagnosed VS for 5 months following a severe traumatic brain injury, fMRI measurements provided evidence that the patient was able to activate very specific brain regions in response to complex commands when asked to carry out imaginary actions. As shown in figure 78.2, when asked to imagine playing tennis, the patient exhibited elevated brain activity in the supplementary motor areas matching the activity profile of normal subjects; similarly, when asked to imagine walking through the rooms of her house, the patient activated the parahippocampal gyrus, posterior parietal cortex, and lateral premotor cortex, regions activated in the normal control subjects carrying out this task (see figure 78.2 legend). At the time of the fMRI study, bedside examination of the patient showed evidence of only brief visual fixation, a possible transitional sign for evolution into MCS (Giacino et al., 2002) but a finding also consistent with permanent vegetative state in some patients (Jennett, 2002). Another examination 11 months later revealed visual tracking to a mirror, another transitional sign consistent with MCS, but no evidence of object manipulation or behavioral manifestations of command following or other external motor responses.

These imaging findings demonstrate a preservation of cognitive function that the clinical bedside examination in this particular patient failed to reveal and indicate a functional level at least consistent with MCS. Nonetheless, the demonstration of command following alone cannot clarify this patient's potential level of conscious awareness and cognitive capacity and leaves open the possibility that higher levels of function could be present. The findings, however, do demonstrate that the fMRI technique can be operationally exchangeable with behavioral evidence of command following as judged clinically at the bedside and in principle could provide a method for diagnosis of a nonbehavioral

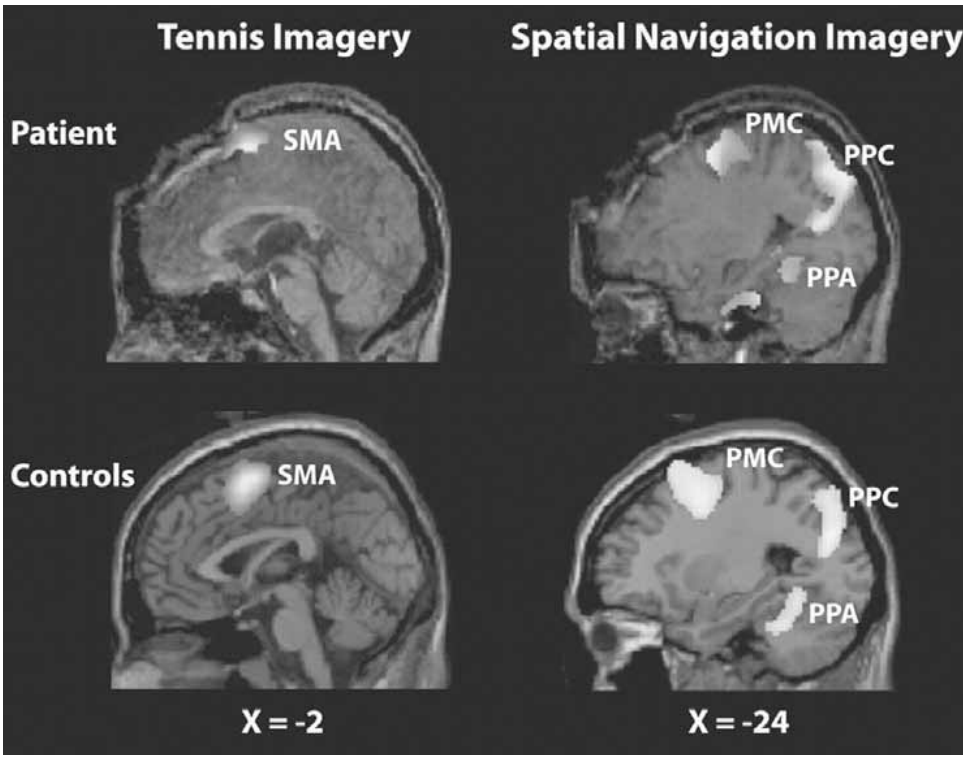


FIGURE 78.2 Command following in vegetative state (Owen et al., 2006). A 23-year-old woman with clinical exam consistent with VS, five months after severe traumatic brain injury with only brief periods of visual fixation, was asked to imagine playing tennis or walking throughout her own house. The regionally selective brain activation patterns obtained from functional magnetic resonance imaging measurements for each condition were identical to those of normal controls. (Reproduced with permission.) (See color plate 91.)

minimally conscious state (Fins & Schiff, 2006). Owen and Coleman (2008) have argued that "a 'non-behavioural fully conscious state' is equally plausible" for this patient, perhaps inferring the capacity for communication and higher cognition as implicit in the complex imaginal aspects of the commands followed. As noted previously, however, MCS patients who show consistent visible motor responses to complex commands and yet are unable to sustain interactive communication are not uncommon (Giacino & Whyte, 2005; Schiff et al., 2005).

More importantly, fMRI findings (or those of any other neuroimaging/measurement technique) should be not considered in isolation of the pathophysiological mechanisms underlying a patient's brain injury. In the case of the patient studied by Owen and colleagues (2006), at least two factors raise doubts about the claim that the patient is likely to be "fully conscious." The severe diffuse axonal injury suffered by the patient and the evident impact of the patient's head injury on frontal systems (seen in the collapsed regions of skull across the patient's frontal lobe visible in both panels of figure 78.2) likely index a significant underlying functional impairment of the frontal executive motor control systems. The cortico-striatopallidal-thalamocortical loop frontal systems are selectively vulnerable at the "circuit" level in the setting of many types of multifocal brain injury (see the following discussion), and metabolic depression of activity in these networks specifically grades with severity of behavioral impairment following diffuse axonal injury (Kato et al., 2007). Importantly, Owen and colleagues demonstrated the integrity of the motor pathways in their patient using transcranial magnetic stimulation methods, ruling out a contribution of interruption of the outflow from the motor cortex to the skeletal muscles accounting for their lack of initiated movements (Owen et al.). Thus a more parsimonious inference is that although this patient could follow commands, they likely remained unable to communicate and carry out goal-directed intentional behaviors because of generalized cognitive impairment.

The findings of the Owen and colleagues (2006) study extend earlier observations of widely preserved language responsive networks in some MCS patients (Schiff et al., 2005) by unambiguously demonstrating higher levels of cognitive function (command following) in a patient with only limited behavioral evidence of nonreflexive movements (limited to brief fixation that gradually improved to show isolated visual tracking). Importantly, such evidence of higher-level cerebral integrative activity indicates a potential substrate for further recovery. As noted earlier, time to recovery from MCS is variable, and some patients can make substantial improvement recovering spontaneous

fluent speech and reliable communication after a year or more of remaining at the MCS level of observable behaviors.

Association between functional recovery and white matter structural changes

Perhaps the most extreme case of late recovery is that of Terry Wallis, a man who at age 40 recovered full expressive and receptive language after remaining in MCS for 19 years following a severe traumatic brain injury suffered in a motor vehicle accident (Schiff & Fins, 2003). Structural magnetic resonance imaging studies of Wallis's brain revealed extensive cerebral and subcortical atrophy particularly affecting the brain stem and frontal lobes. Diffusion tensor magnetic resonance imaging (DTI), a technique that quantifies proton diffusion (in this case primarily water) and is used to measure the restriction of the direction of water flow (fractional anisotropy) as a proxy for axonal fiber integrity, revealed marked reduction in fractional anisotropy consistent with severe diffuse axonal injury. The severity of loss of axonal fibers can be appreciated qualitatively in figure 78.3 and quantitatively in figure 78.4 by comparing the volumes of the medial corpus callosum in 20 normal subjects to the measured values from Wallis's brain.

In a longitudinal DTI study of Wallis's brain, Voss and colleagues (2006) identified that in contrast to the overall severe reduction of brain connectivity demonstrated by reduced fractional anisotropy in most brain regions, DTI measurements also revealed large regions of increased fractional anisotropy in posterior brain white matter not seen in the normal subjects (figures 78.3 and 78.4) that had a prominent left-right directionality (colored red in the figure). These regions of posterior white matter anisotropy showed less strong directionality when measured in a second DTI study 18 months later, whereas similar significant increases in fractional anisotropy and left-right directionality were identified within the midline cerebellar white matter. At the time of the second study Wallis showed significant clinical improvements in motor control (limited use of lower extremities and left upper extremity, and improvement in dysarthria) over the intervening time period between the two studies correlating with the changes observed in the cerebellum (figures 78.3 and 78.4). These findings suggest the possibility that structural changes within Wallis's white matter may have played a role in his functional recovery of limb, trunk, and articulatory movements. A prospective study of severely brain-injured patients followed from early injury until approximately one year after injury found similar changes in recovery of initially reduced fractional anisotropy

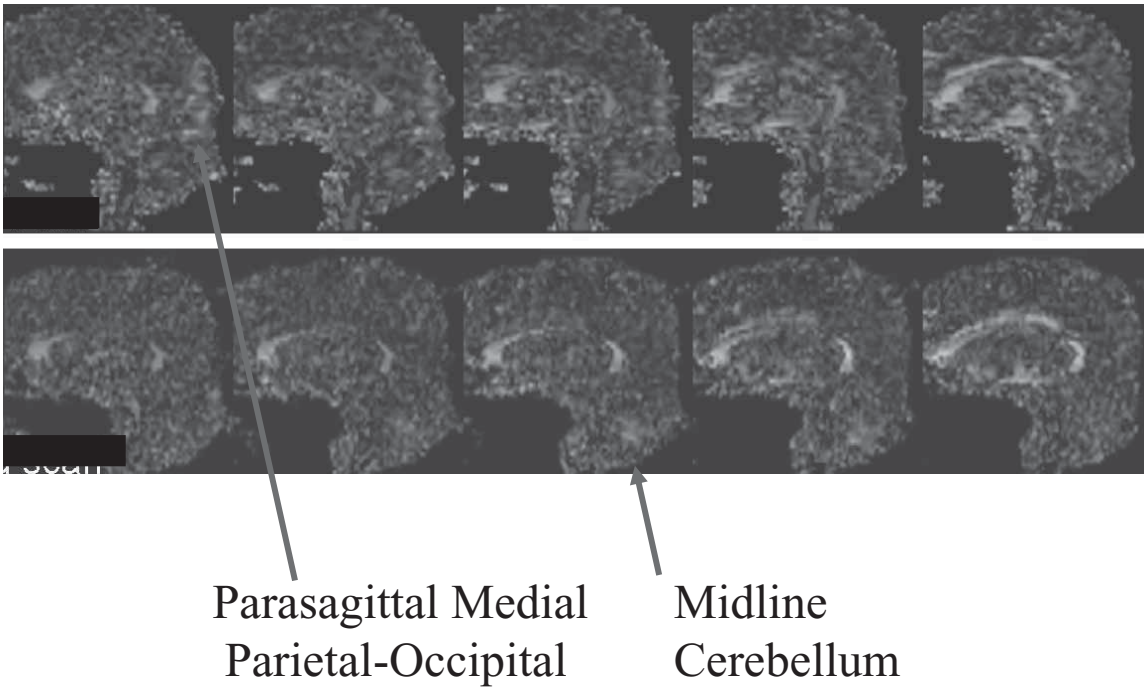


FIGURE 78.3 Diffusion tensor imaging studies of patient with late recovery (19 years) from MCS (Voss et al., 2006). Fractional anisotropy maps showing fiber tracks: red, fibers with left-right directionality; blue, fibers with up-down directionality; green, fibers with anterior-posterior directionality. Top images show volume loss of the corpus callosum throughout the medial component and regions in parieto-occipital white matter with prominent left-right directionality. Bottom row images show fractional anisotropy maps obtained 18 months later that demonstrate reduction of left-right direction in parieto-occipital regions with increased anisotropy noted in the midline cerebellum. (See color plate 92.)

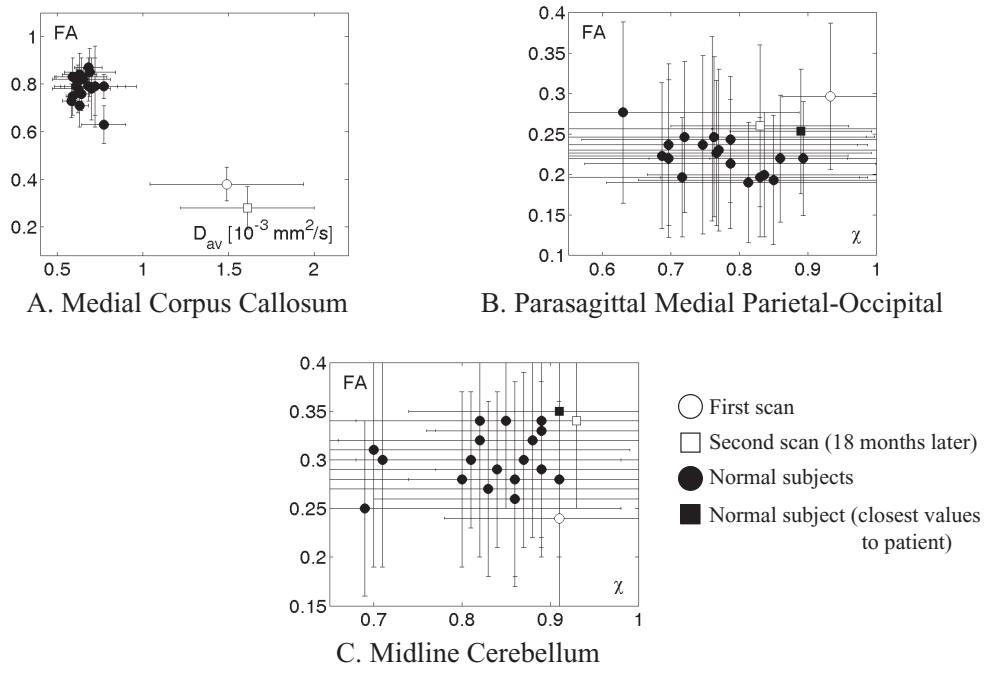


FIGURE 78.4 Quantitive comparison of midline cerebellum fractional anisotropy (FA) versus apparent diffusion (D_{av}) and left-right directionality (χ) for regions of interest following late recovery from MCS. Open-circle values were obtained from patient's first scan; open-square values were obtained from second scan; filled-circle values are from 20 normal subjects; filled squares indicate values of individual normal subject closest to patient value. (*A*) Medial corpus callosum shows marked loss of fractional anisotropy in patient measured on both scans. (*B*) Parasagittal medial parietal-occipital region shows strong left-right directionality and increased fractional anisotropy in patient's first scan (see Voss et al., 2006, for details). (*C*) Midline cerebellum shows marked increase in fractional anisotropy over two studies. (Plots from Voss et al., 2006, with permission.)

in the internal capsule and centrum ovale, which later showed normal or supernormal levels of fractional anisotropy parallel to the axis of axonal fibers in the patients who recovered neurological function (Sidaros et al., 2008).

These findings can be compared with experimental studies that demonstrate late remodeling of cortical white matter connections following ischemic (stroke) injuries in monkeys. Following strokes, existing neurons are found to form extensive axonal arborizations to establish novel connections in intact brain regions (Dancause et al., 2005). Taken together, these observations suggest that the normal recovery process includes a component of structural remodeling that may stall and potentially be restarted under the appropriate conditions. Whether this process may have restarted in Wallis at close to the time of his reemergence of spoken language is not known. However, approximately 18 months prior to speaking he had been placed on paroxetine (a serotonin reuptake inhibitor), and his behavioral responsiveness for several years after emergence from MCS remained delicately dependent on the level of this medicine (unpublished observations). Recent studies suggest that chronic exposure to serotonin reuptake inhibitors may promote some aspects of neuronal plasticity (Wang, David, Monckton, Battaglia, & Hen, 2008).

Circuit mechanisms underlying forebrain dysfunction follow severe brain injury

The preceding observations that patients with no apparent, or only very limited, evidence of awareness may retain or regain large-scale integrative brain activity raises the critical question of what mechanisms might underlie their failure to exhibit goal-directed behavior despite, in some cases, the integrity of motor pathways. An important clue is given by recent studies of the anatomic pathology associated with severe brain injuries producing chronic disorders of consciousness. Autopsy studies of both traumatic and nontraumatic severe brain injuries demonstrate a common finding in permanent vegetative state of widespread death of thalamic neurons in patients with either multifocal cerebral neuronal death from anoxia or diffuse axonal injury producing multiple disconnections of fiber pathways (Adams, Graham, & Jennett, 2000). The severe bilateral thalamic damage seen in both types of injury is not invariably associated with diffuse cortical damage. This finding can be understood in the context of the dense innervation of the thalamus by the cortex so that multifocal loss of cortical neurons is reflected by a greater impact on thalamic neurons that undergo neuronal death following sufficient deafferentation.

Importantly, specific components of the thalamus are more sensitive reporters of global, multifocal injuries. The nuclei that comprise the central thalamus (the intralaminar nuclei and related paralaminar nuclei) show the most marked cell loss with traumatic injuries and the degree of neuronal loss observed in these nuclei grades with outcome (Maxwell, MacKinnon, Smith, McIntosh, & Graham, 2006). Loss of neurons within the anterior intralaminar nuclei (central lateral nucleus, central medial, paracentralis) is identified in patients who recovered to a level of moderate disability with progressive involvement of more ventral and lateral nuclei of the central thalamus (posterior intralaminar group) seen in patients with severe disability and permanent vegetative state. Figure 78.5 diagrams the pattern of structural injury to the central thalamus and midbrain seen in patients with bilateral focal strokes that produce immediate coma and enduring vegetative state or minimally conscious state (Castaigne et al., 1981; Schiff & Plum, 2000). The neuronal populations overlap those neurons that undergo progressive deafferentation with increasingly severe multifocal brain injuries (Maxwell et al.) as shown in figure 78.5*B*, whereas the first loss of neurons associated with moderate disability following brain injury occurs in the more anterior intralaminar regions.

The observation that neurons within the central thalamus are more vulnerable to multifocal injuries can be first understood simply in terms of their geometry of connections: they have wide point-to-point connectivity across the four lobes of the cerebral hemisphere and are positioned to integrate neuronal cell death across these large territories (van der Werf, Witter, & Groenewegen, 2002; Scannell, Burns, Hilgetag, O'Neil, & Young, 1999). The marked impact of isolated injuries to these neuronal populations noted previously indicates that neurons in the central thalamus also have a specific causal role in producing impaired integrative forebrain function when damaged directly (Castaigne et al., 1981; Schiff & Plum, 2000). This contribution to forebrain dysfunction following either focal injury or deafferentation as a consequence of diffuse brain injuries is most likely linked to their established contribution to normal mechanisms of arousal regulation (Schiff, 2008).

Neurons within the central thalamus have anatomical and physiological specializations that make them uniquely positioned to control arousal levels and the dynamic

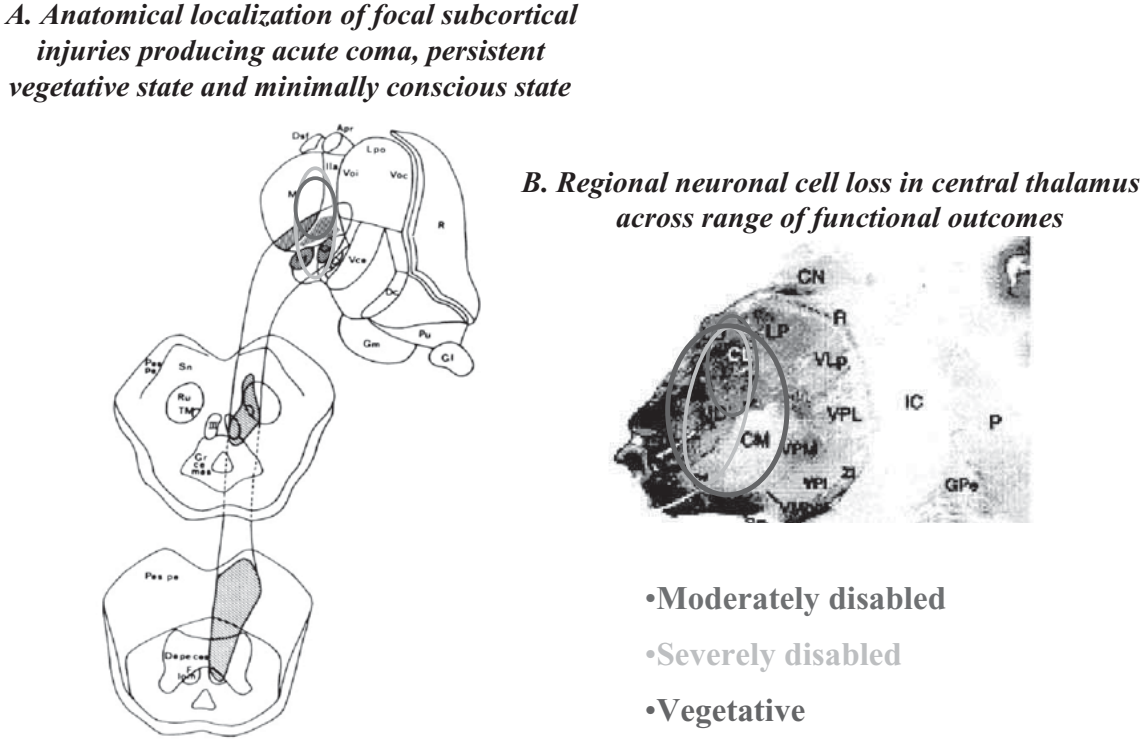


Figure 78.5 Contributions of the central thalamus to disorders of consciousness. (*A*) Focal injury patterns in the central thalamus associated with coma, vegetative state, and minimally conscious state. (*B*) Regional neuronal cell loss in central thalamus across the range of function outcomes. Moderately disabled (red): neuronal loss in median dorsalis, rostral central medial, central lateral and paracentral nuclei. Severely disabled (green): includes moderately disabled regions plus neuronal loss from median dorsalis, caudal central medial, parafasicular nucleus. Permanent vegetative state (blue): all of above and centromedian nucleus. (Figure elements adapted from Castaigne et al., 1981, and Münkle, Waldvogel, & Faull, 2000.) (See color plate 93.)

patterns of activation across large-scale cerebral networks (Groenewegen & Berendse, 1994; Llinas, Leznik, & Urbano, 2002; Jones, 2001; van der Werf et al., 2002; Purpura & Schiff, 1997; Schiff & Purpura, 2002; Schiff, 2008). The central thalamus is interposed between brain stem/basal forebrain "arousal systems" that control overall levels of depolarization of cortical and thalamic neurons and frontal "executive systems" that organize premotor shifts of attention and adjust the level of alertness and cognitive effort (Kinomura, Larsson, Gulyás, & Roland, 1996; Paus et al., 1997; van der Werf et al., 2002; reviewed in Schiff, 2008). In addition, individual neurons within these cellular aggregates are targets of a variety of dedicated brain stem sensory relays that have evolved to quickly capture attention and redirect behavior. The neurons within the central thalamus are recruited in response to increasing cognitive demands, stress, fatigue, and other perturbations that reduce behavioral performance (Kinomura et al., 1996; Paus et al., 1997; Nagai et al., 2004; Schiff, Hudson, & Purpura, 2002). From a functional point of view, central thalamic neurons may act as gain controls for distributed networks across the cortex, thalamus, and basal ganglia through a common mechanism of shifting the overall level of balanced excitatory and inhibitory synaptic barrage associated with an "UP state" like phenomena during wakefulness (Rudolph, Pelletier, Paré, & Destexhe, 2005; Shu, Hasenstaub, & McCormick, 2003; Schiff, 2008). Adjustments of firing rates within the central thalamus by mesial frontal control systems could thus depolarize neurons across the cerebral cortex and striatum and selectively gate their activation.

As illustrated in figure 78.6, an important aspect of vulnerability of the forebrain to dysfunction produced by deafferentation of the central thalamus is the key role of neurons from the intralaminar nuclei (both central lateral nucleus and parafasicularis nucleus) that project to the medium spiny neuron (MSN) of the striatum (Lacey, Bolam, & Magill, 2007). The MSNs have a "high-threshold"

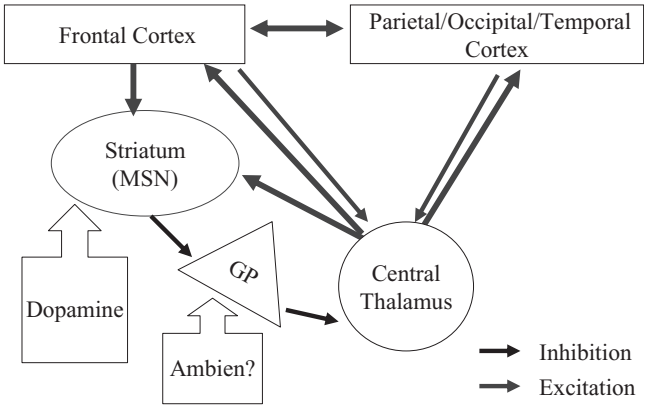


FIGURE 78.6 Connections of the central thalamus underlying their role in forebrain arousal regulation. See text.

UP state making it difficult to bring these neurons to their firing threshold; they require both sufficient levels of dopamine neuromodulation and a high level of spontaneous synaptic activity arising from excitatory corticostriatal and thalamostriatal inputs (Grillner et al., 2005). The output from the MSNs, in turn, opposes a tonic inhibitory activity of the globus pallidus interna on the thalamus. Loss of input from the central thalamus can potentially shut down the MSNs both through withdrawal of direct excitatory striatal projections (Lacey, Bolam, & Magill, 2007) and through downregulation of the frontocortical regions that provide the main corticostriatal input (Schiff & Posner, 2007). The circuit model shown in figure 78.6 provides an explanation for well-known observations that some patients with either mesial frontal or thalamic/midbrain injuries may respond to dopaminergic and similar agents that facilitate output of the MSNs and mesial frontal systems (Whyte et al., 2005). In addition, this simplified circuit description predicts that direct activation of the central thalamus would reverse downregulation across the corticostriatalpallidal-thalamocortical system produced by severe brain injury. It also offers an explanation of a surprising, paradoxical phenomenon recently described that zolpidem (the sedative agent known as Ambien that is a nonbenzodiazepine hypnotic that potentiates GABA-A receptors expressed in large quantity in the globus pallidus interna) can improve alertness and behavioral responsiveness in some severely brain-injured patients (Breufel-Courbon et al., 2007; Schiff & Posner, 2007).

Figure 78.7 shows unpublished positron emission tomography (PET) studies of resting brain metabolism from a patient who remained in MCS for two years prior to speaking after receiving a single dose of zolpidem. Over a five-year period with repeated dosing the patient gradually improved to emergence from MCS and reliable communication on or off the medication. During the patient's off-drug periods, however, he was cognitively slowed and unable to complete oral feeding. As seen in figure 78.7, in the off-medication state a marked downregulation of metabolic activity is apparent across the corticostriatalpallidal-thalamocortical system of the anterior forebrain, which reverses with application of zolpidem. These findings reproduce identical shifts in forebrain metabolism measured by PET in another published study of zolpidem-induced recovery of function in a severely brain-injured patient (Breufel-Courbon et al., 2007). As noted in figure 78.6, because zolpidem binds selectively to the alpha-1 subunit of the GABA-A receptor that is highly expressed in the globus pallidus interna (GPi), it may act to suppress disinhibited GPi neurons facilitating thalamocortical and thalamostriatal afferents outflow to restore sufficient background synaptic activity to drive the MSNs and restablish the long-loop network activation seen in the ON state (Schiff & Posner, 2007).

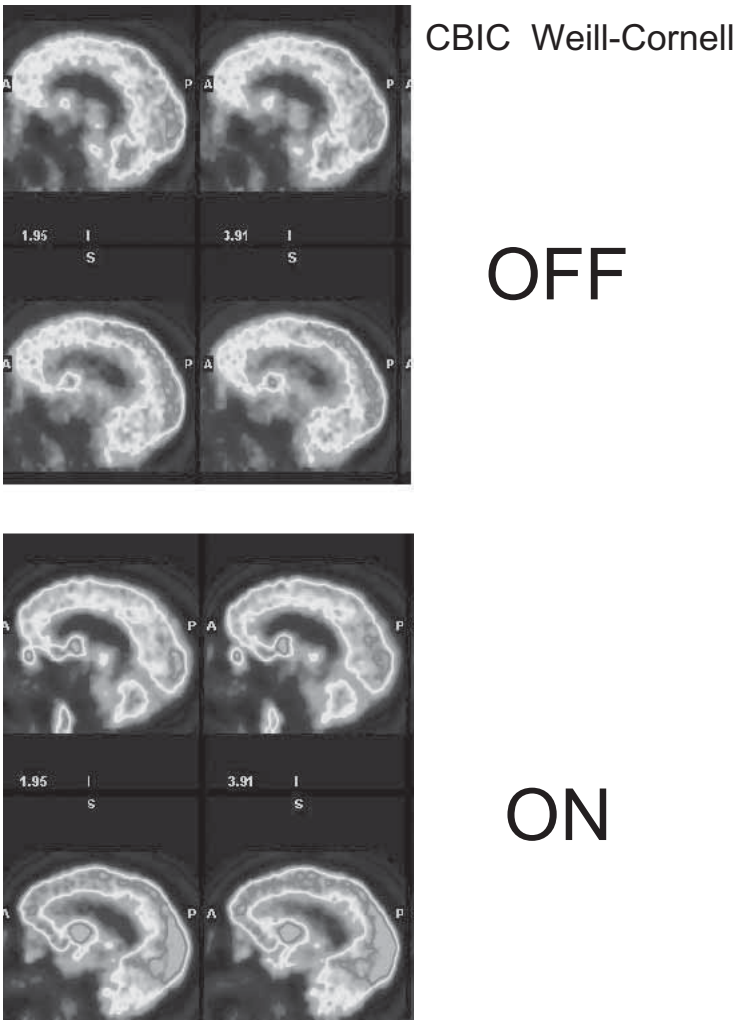


FIGURE 78.7 Changes in cerebral metabolism associated with zolpidem administration in severe brain injury. Two sets of parasagittal images of fluorodeoxyglucose positron emission tomography studies obtained in an awake, severely brain-injured patient are shown obtained one day apart. The OFF images show the resting metabolic profile in an awake state and demonstrate qualitative downregulation of metabolism in the anterior forebrain (frontal, prefrontal cortex), basal ganglia, and thalamus. The ON images show the resting metabolic profile in an awake state 45 minutes after administration of the drug zolpidem. A marked qualitative increase in metabolism is observed in the anterior forebrain, basal ganglia, and thalamus. In addition, overall metabolic activity across cerebral structures is increased. (See color plate 94.)

Modulating behavioral responsiveness in the severely injured brain using central thalamic deep-brain stimulation

In the context of the model diagrammed in figure 78.6, the observations of indirect control of central thalamic output by facilitating MSNs firing through dopaminergic modulation or withdrawal of GPi inhibition predict that directly controlling excitatory output from the central thalamus would produce more consistent and sustained effects. Direct activation of central thalamic neurons through deep-brain electrical stimulation (DBS) has been proposed as an experimental therapeutic strategy (Schiff, Plum, & Rezai, 2000, 2002; Schiff & Purpura, 2002). A single-subject study of central thalamic DBS tested the underlying hypothesis that electrical stimulation could restore arousal regulation and promote greater behavioral responsiveness in a 38-year-old man who had remained in MCS for six years (Schiff et al., 2007). Although the patient was unable to communicate reliably, earlier neuroimaging studies showed preservation of a large-scale cerebral language network (Schiff et al., 2005), suggesting that a substrate for further recovery might exist. Central thalamic DBS was proposed to, in part, substitute for the loss of top-down monitoring and regulation of arousal level by mesial frontal cortical regions and ascending brain stem arousal system inputs damaged by injury and functionally impaired through the same general circuit-level disturbance as diagrammed in figure 78.6 (multifocal cerebral

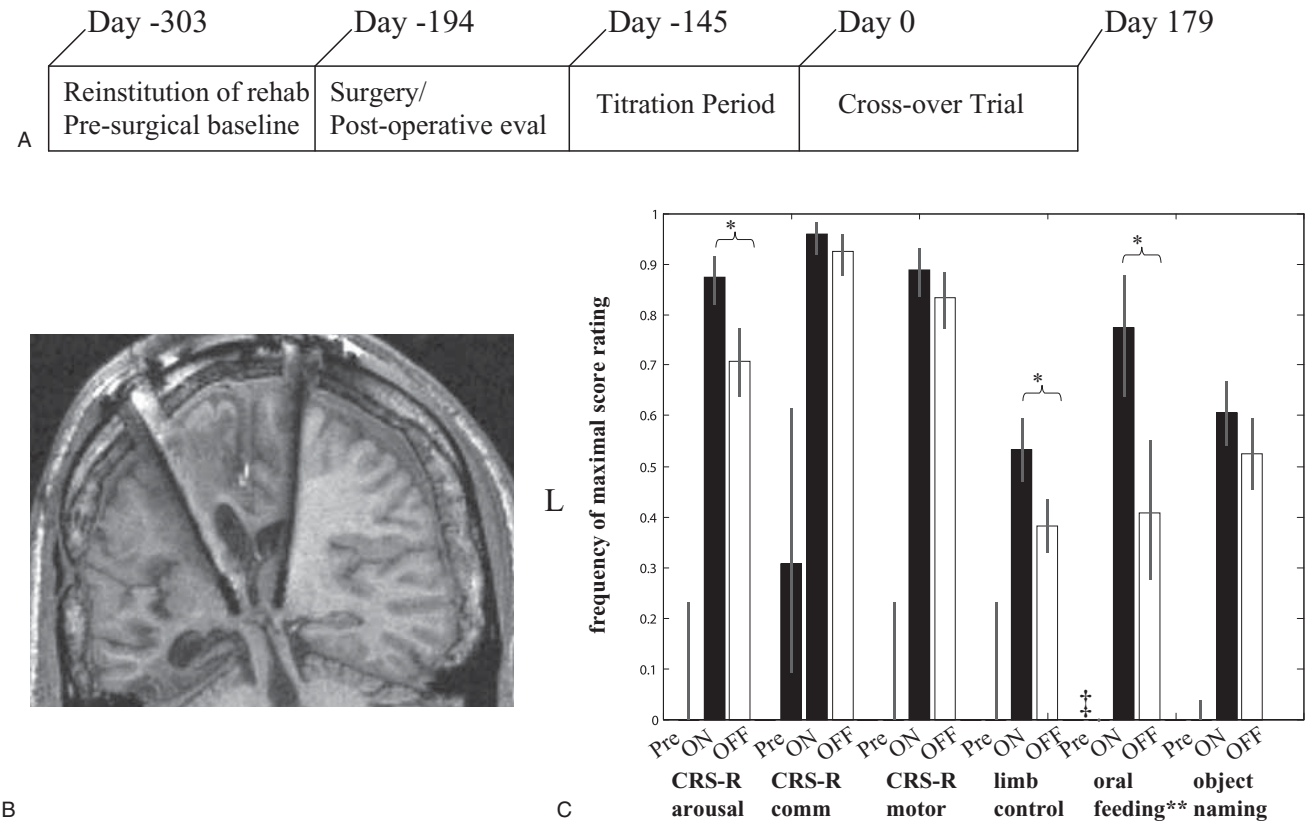


FIGURE 78.8 Central thalamic electrical stimulation in the minimally conscious state. (*A*) Time line of a single-subject study of deep-brain stimulation (DBS). (*B*) Placement of electrodes in the central thalamus. (*C*) Results of 180-day crossover study of DBS. Statistically significant improvements during the crossover study in attentive behavior, oral feeding, and limb control are marked with asterisks (see text and Schiff et al., 2007, for further details).

deafferentation and dampening of forebrain arousal regulation within the mesial frontal and central thalamic network connections).

Figure 78.8 displays the overall design of the study and main results (Schiff et al., 2007). The patient was monitored over an 11-month period that included six months of behavioral baseline evaluations (four months prior to surgery, two months following surgery prior to titration testing of stimulation parameters) that indicated no change in level of behavioral responsiveness as compared to testing done prior to discharge to chronic care facility more than two years before the start of the trial. Electrodes were placed bilaterally in the central thalamus targeting the intralaminar nuclei with the central lateral nucleus in midposition of the electrode contacts (see figure 78.8). A five-month period of titration testing demonstrated immediate and accumulating effects of DBS that included the emergence of consistent and intelligible spoken language, recovery of limb control, and the capacity for oral feeding (the patient was fed by a gastrotomy tube for five years prior to this phase of the study). Following titration testing, a single stimulation parameter set was chosen for testing in a blinded six-month, 30-day alternating ON-versus-OFF study. The results of the crossover study are shown in the graph in figure 78.8*C*. The crossover study demonstrated a robust overall effect on behavioral responsiveness measured by a formal quantitative assessment tool (see figure 78.8 legend) and significant stimulation-dependent modulation of oral feeding and limb control. Notably, all functional testing showed marked improvement against the six-month prestimulation baselines, but several measures hit a ceiling of improvement whether DBS was on or off during this later phase of the trial.

The observed carryover effects of improvements from the ON to the OFF state may be a result of strengthening activated synapses or other mechanisms underlying neuronal plasticity. These findings can be compared to evidence of accumulating effects of central thalamic electrical stimulation as shown in a rodent model. Herrera and colleagues studied rats undergoing continuous unilateral electrical stimulation of the central lateral nucleus (Shirvalkar, Seth, Schiff, & Herrera, 2006). The stimulated rats exhibited enhancements of observed improvements in an untrained goal-directed seeking behavior that requires object recognition memory over three days of testing (figure 78.9). The study also found increases in exploratory motor behaviors and grooming activity (Shirvalkar et al.). Separate experiments

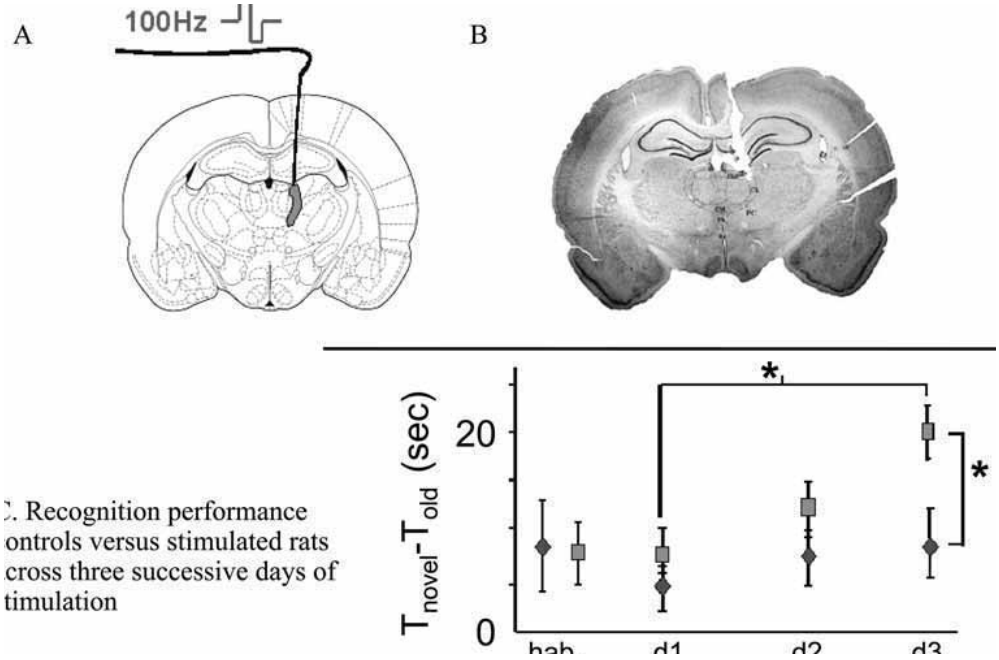


FIGURE 78.9 Facilitation of recognition memory with continuous electrical stimulation of the rat central thalamus (Shirvalkar, Seth, Schiff, & Herrera, 2006). (*A*) Electrode placement in the central thalamus. (*B*) Representative gross histological section. (*C*) Comparison of sham-stimulated controls and stimulated rat cohorts on object recognition task across three successive days of stimulation. Stimulated rats show increased performance and accumulation of effects. (See color plate 95.)

evaluated cerebral gene expression following exposure to the same electrical stimulation parameters. Central thalamic DBS produced upregulation of memory-related immediate early genes in the anterior cingulate cortex, motor cortex, and hippocampus indicating a link between DBS and cellular mechanisms of memory formation (figure 78.10). The pattern of memory-related immediate early gene activation observed shows a variation across cortical layers consistent with the known innervation pattern of the central lateral intralaminar afferents and laminar activation profile with electrical stimulation (figure 78.10; Llinas et al., 2002). Of note, similar changes in neocortical gene expression are observed after periods of sleep following induction of long-term potentiation (Ribeiro et al., 2002). These findings suggest that changes in cellular gene expression may be part of the mechanism for the observed carryover effects in the human trial and that effects on the sleep-wake periods may also have a role.

Conclusions and future directions

The evolving understanding of the recovery of consciousness in the severely injured brain described in this chapter indicates the need for developing new tools to guide longitudinal assessments of brain function. In addition, systematic efforts to evaluate time trials of pharmacologic agents and novel therapeutics should become part of the general approach to the problem. The goal of such work should be to formulate more precise and testable models at the circuit level that are predictive of patterns of recovery and response to treatment. An important aspect of recovery is the evidence that plastic changes can arise very late after injury and likely depend on interactions between circuit mechanisms that reestablish goal-directed behavior initiating changes in the efficacy, or simply engagement of, cellular repair mechanisms. The findings reviewed indicate that impaired forebrain dynamics may in some patients be dramatically altered by pharmacologic agents or central thalamic electrical stimulation. That carryover effects are seen with brain stimulation suggests that this induced recovery process may share common mechanisms with the slow changes observed in the Wallis case; however, the biological mechanisms remain unclear.

A unifying speculation is that changes in the overall level of activity within the circuit diagrammed in figure 78.6 occur in the setting of spontaneous or induced recovery of goal-directed behavior and communication that are the hallmarks of observed recovery of consciousness—although this process clearly can also evolve without correlative observable behavior! The main variable controlling the recovery of consciousness may be gradual improvements in the integrity of basic arousal regulation processes within the corticothalamic (corticostriatopallidal-thalamocortical) system. If this recovery process can be supervised, we will be able to understand in much greater detail how brain mechanisms link integrative dynamics of circuit-level functions such as attention, working memory, and motor preparation to the

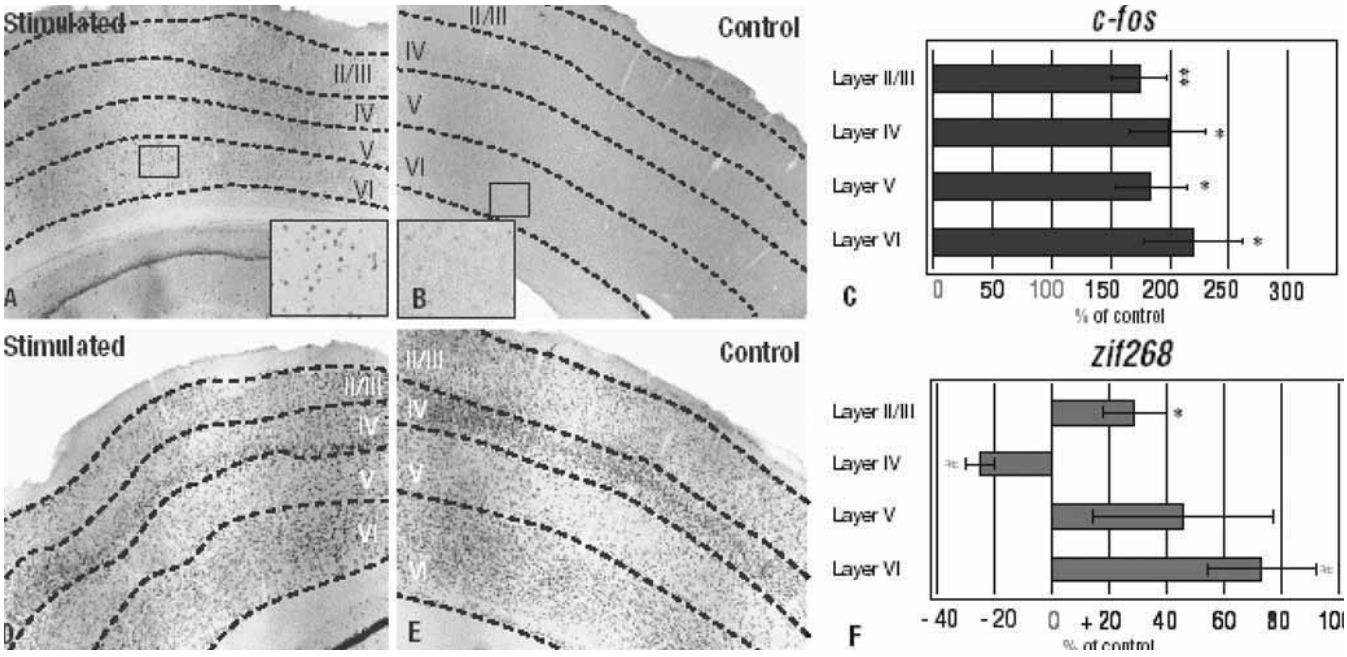


FIGURE 78.10 Gene expression changes in the motor cortex following central thalamic electrical stimulation. Gene expression changes associated with 30 minutes of electrical stimulation of the central thalamus in two immediate early genes, *c-fos* and *zif268*, are shown (Shirvalkar et al., 2006). A broad increase of *c-fos* across cortical lamina is seen, consistent with increased synaptic activity. A laminar specific pattern of changes in *zif268*, a memory-related gene, may link to accumulation effects seen in figure 78.9 and possibly to carryover effects observed in human-subject DBS study shown in figure 78.8 (see text and Shirvalkar et al., 2006, for further details). (See color plate 96.)

cellular learning and memory mechanisms that support continuing functional improvements. Viewed from this perspective, controlling the interplay of observed behavioral and measured brain recovery may be the overarching goal of a cognitive neuroscience of the recovery of consciousness.

ACKNOWLEDGMENTS The author acknowledges the support of the NIH-NINDS, NICHD, Charles A. Dana Foundation, James S. McDonnell Foundation, and Jerold B. Katz Foundation and the contributions of S. Shah and S. Williams for their careful and helpful review.

REFERENCES

ADAMS, J. H., GRAHAM, D. I., & JENNETT, B. (2000). The neuropathology of the vegetative state after acute insult. *Brain*, *123*, 1327–1338.

BREFEL-COURBON, C., PAYOUX, P., ORY, F., SOMMET, A., SLAOUI, T., RABOYEAU, G., LEMESLE, B., PUEL, M., MONTASTRUC, J. L., DEMONET, J. F., & CARDEBAT, D. (2007). Clinical and imaging evidence of zolpidem effect in hypoxic encephalopathy. *Ann. Neurol.*, *62*(1), 102–105.

CASTAIGNE, P., LHERMITTE, F., BUGE, A., ESCOUROLLE, R., HAUW, J. J., & LYON-CAEN, O. (1981). Paramedian thalamic and midbrain infarcts: Clinical and neuropathological study. *Ann. Neurol.*, *10*(2), 127–148.

COLEMAN, M. R., RODD, J. M., DAVIS, M. R., JOHNSRUDE, I. S., MENON, D. K., PICKARD, J. D., & OWEN, A. M. (2007). Do vegetative patients retain aspects of language comprehension? Evidence from fMRI. *Brain*, *130*(Pt. 10), 2494–2507.

DANCAUSE, N., BARBAY, S., FROST, S. B., PLAUTZ, E. J., CHEN, D., ZOUBINA, E. V., STOWE, A. V., & NUDO, R. J. (2005). Extensive cortical rewiring after brain injury. *J. Neurosci.*, *25*, 10167–10179.

FINS, J. J., & SCHIFF, N. D. (2006). Shades of gray: New insights into the vegetative state. *Hastings Cent. Rep.*, *36*(6), 8.

GIACINO, J. T., ASHWAL, S., CHILDS, N., CRANFORD, R., JENNETT, B., KATZ, D. I., KELLY, J. P., ROSENBERG, J. H., WHYTE, J., ZAFONTE, R. D., & ZASLER, N. D. (2002). The minimally conscious state: Definition and diagnostic criteria. *Neurology*, *58*, 349–353.

GIACINO, J. T., & WHYTE, J. (2005). The vegetative state and minimally conscious state: Current knowledge and remaining questions. *J. Head Trauma Rehabil.*, *20*(1), 30–50.

GRILLNER, S., HELLGREN, J., MÉNARD, A., SAITOH, K., & WIKSTROM, M. A. (2005). Mechanisms for selection of basic motor programs—Roles for the striatum and pallidum. *Trends Neurosci.*, *28*(7), 364–370.

GROENEWEGEN, H., & BERENDSE, H. (1994). The specificity of the "nonspecific" midline and intralaminar thalamic nuclei. *Trends Neurosci.*, *17*, 52–66.

JENNETT, B. (2002). *The vegetative state*. Cambridge, UK: Cambridge University Press.

JONES, E. G. (2001). The thalamic matrix and thalamocortical synchrony. *Trends Neurosci.*, *24*, 595–601.

KATO, T., NAKAYAMA, N., YASOKAWA, Y., OKUMURA, A., SHINODA, J., & IWAMA, T. (2007). Statistical image analysis of cerebral glucose metabolism in patients with cognitive impairment

following diffuse traumatic brain injury. *J Neurotrauma*, *24*(6), 919–926.

Kinomura, S., Larsson, J., Gulyás, B., & Roland, P. E. (1996). Activation by attention of the human reticular formation and thalamic intralaminar nuclei. *Science*, *271*, 512–515.

Kobylarz, E. J., & Schiff, N. D. (2005). Neurophysiological correlates of persistent vegetative and minimally conscious states. *Neuropsychol. Rehabil.*, *15*, 323–332.

Kübler, A., & Kotchoubey, B. (2007). Brain-computer interfaces in the continuum of consciousness. *Curr. Opin. Neurol.*, *20*(6), 643–649.

Lacey, C. J., Bolam, J. P., & Magill, P. J. (2007). Novel and distinct operational principles of intralaminar thalamic neurons and their striatal projections. *J. Neurosci.*, *27*(16), 4374–4384.

Lammi, M. H., Smith, V. H., Tate, R. L., & Taylor, C. M. (2005). The minimally conscious state and recovery potential: A follow-up study 2 to 5 years after traumatic brain injury. *Arch. Phys. Med. Rehabil.*, *86*(4), 746–754.

Laureys, S., Pellas, F., Van Eeckhout, P., Ghorbel, S., Schnakers, C., Perrin, F., Berré, J., Faymonville, M. E., Pantke, K. H., Damas, F., Lamy, M., Moonen, G., & Goldman, S. (2005). The locked-in syndrome: What is it like to be conscious but paralyzed and voiceless? *Prog. Brain Res.*, *150*, 495–511.

Llinas, R. R., Leznik, E., & Urbano, F. J. (2002). Temporal binding via cortical coincidence detection of specific and nonspecific thalamocortical inputs: A voltage-dependent dye-imaging study in mouse brain slices. *Proc. Natl. Acad. Sci. USA*, *99*, 449–454.

Maxwell, W. L., MacKinnon, M. A., Smith, D. H., McIntosh, T. K., & Graham, D. I. (2006). Thalamic nuclei after human blunt head injury. *J. Neuropathol. Exp. Neurol.*, *65*(5), 478–488.

Münkle, M. C., Waldvogel, H. J., & Faull, R. L. (2000). The distribution of calbindin, calretinin and parvalbumin immunoreactivity in the human thalamus. *J. Chem. Neuroanat.*, *19*, 155–173.

Nagai, Y., Critchley, H. D., Featherstone, E., Fenwick, P. B. C., Trimble, M. R., & Dolan, R. J. (2004). Brain activity relating to the contingent negative variation: An fMRI investigation. *NeuroImage*, *21*(4), 1232–1241.

Owen, A. M., & Coleman, M. R. (2008). Functional neuroimaging of the vegetative state. *Nat. Rev. Neurosci.*, *9*(3), 235–243.

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science*, *313*(5792), 1402.

Paus, T., Zatorre, R., Hofle, N., Caramanos, Z., Gotman, J., Petrides, M., & Evans, A. (1997). Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task. *J. Cogn. Neurosci.*, *9*, 392–408.

Posner, J., Saper, C., Schiff, N., & Plum, F. (2007). *Plum and Posner's diagnosis of stupor and coma* (4th ed.). Oxford, UK: Oxford University Press.

Purpura, K., & Schiff, N. D. (1997). The thalamic intralaminar nuclei: A role in visual awareness. *Neuroscientist*, *3*, 8–15.

Ribeiro, S., Mello, C. V., Velho, T., Gardner, T. J., Jarvis, E. D., & Pavlides, C. (2002). Induction of hippocampal long-term potentiation during waking leads to increased extrahippocampal zif-268 expression during ensuing rapid-eye-movement sleep. *J. Neurosci.*, *22*(24), 10914–10923.

Rudolph, M., Pelletier, J. G., Paré, D., & Destexhe, A. (2005). Characterization of synaptic conductances and integrative properties during electrically induced EEG-activated states in neocortical neurons in vivo. *J. Neurophysiol.*, *94*(4), 2805–2821.

Scannell, J. W., Burns, G. A., Hilgetag, C. C., O'Neil, M. A., & Young, M. P. (1999). The connectional organization of the cortico-thalamic system of the cat. *Cereb. Cortex*, *9*(3), 277–299.

Schiff, N. D. (2004). The neurology of impaired consciousness: Challenges for cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (3rd ed., pp. 1121–1132). Cambridge, MA: MIT Press.

Schiff, N. D. (2005). Modeling the minimally conscious state: Measurements of brain function and therapeutic possibilities. *Prog. Brain Res.*, *150*, 477–497.

Schiff, N. D. (2006). Multimodal imaging studies of disorders of consciousness. *J. Head Trauma Rehabil.*, *21*(5), 388–397.

Schiff, N. D. (2007). Bringing neuroimaging closer to diagnostic use in the severely injured brain. *Brain*, *130*(Pt. 10), 2482–2483.

Schiff, N. D. (2008). Central thalamic contributions to arousal regulation and neurological disorders of consciousness. *Ann. NY Acad. Sci.*, *1129*, 105–118.

Schiff, N. D., & Fins, J. J. (2003). Hope for "comatose" patients. *Cerebrum*, *5*(4), 7–24.

Schiff, N. D., Giacino, J. T., Kalmar, K., Victor, J. D., Baker, K., Gerber, M., Fritz, B., Eisenberg, B., Biondi, T., O'Connor, J., Kobylarz, E. J., Farris, S., Machado, A., McCagg, C., Plum, F., Fins, J. J., & Rezai, A. R. (2007). Behavioral improvements with thalamic stimulation after severe traumatic brain injury. *Nature*, *448*, 600–603.

Schiff, N. D., Hudson, A. E., & Purpura, K. P. (2002). Modeling wakeful unresponsiveness: Characterization and microstimulation of the central thalamus. *Soc. Neurosci. Abstracts*, *62*, 12.

Schiff, N. D., & Plum, F. (2000). The role of arousal and "gating" systems in the neurology of impaired consciousness. *J. Clin. Neurophysiol.*, *17*, 438–452.

Schiff, N. D., Plum, F., & Rezai, A. R. (2002). Developing prosthetics to treat cognitive disabilities resulting from acquired brain injuries. *Neurol. Res.*, *24*, 116–124.

Schiff, N. D., & Posner, J. P. (2007). Another *Awakenings*. *Ann. Neurol.*, *62*, 5–7.

Schiff, N. D., & Purpura, K. P. (2002). Towards a neurophysiological basis for cognitive neuromodulation through deep brain stimulation. *Thalamus Relat. Syst.*, *2*, 55–69.

Schiff, N. D., Ribary, U., Moreno, D., Beattie, B., Kronberg, E., Blasberg, R., Giacino, J., McCagg, C., Fins, J. J., Llinas, R., & Plum, F. (2002). Residual cerebral activity and behavioral fragments in the persistently vegetative brain. *Brain, 125*(6), 1210–1234.

Schiff, N. D., Rodriguez-Moreno, D., Kamal, A., Kim, K. H., Giacino, J. T., Plum, F., & Hirsch, J. (2005). fMRI reveals large-scale network activation in minimally conscious patients. *Neurology*, *64*, 514–523.

Shirvalkar, P., Seth, M., Schiff, N. D., & Herrera, D. G. (2006). Cognitive enhancement with central thalamic electrical stimulation. *Proc. Natl. Acad. Sci. USA*, *103*(45), 17007–17012.

Shu, Y., Hasenstaub, A., & McCormick, D. A. (2003). Turning on and off recurrent balanced cortical activity. *Nature.*, *423*, 288–293.

Sidaros, A., Engberg, A. W., Sidaros, K., Liptrot, M. G., Herning, M., Petersen, P., Paulson, O. B., Jernigan, T. L., & Rostrup, E. (2008). Diffusion tensor imaging during recovery from severe traumatic brain injury and relation to clinical outcome: A longitudinal study. *Brain.*, *131*, 559–572.

Van der Werf, Y. D., Witter, M. P., & Groenewegen, H. J. (2002). The intralaminar and midline nuclei of the thalamus: Anatomical and functional evidence for participation in processes of arousal and awareness. *Brain Res. Brain Res. Rev.*, *39*(2–3), 107–140.

von Domburg, P. H., ten Donkelaar, H. J., & Notermans, S. L. (1996). Akinetic mutism with bithalamic infarction: Neurophysiological correlates. *J. Neurol. Sci.*, *139*, 58–65.

Voss, H. U., Uluç, A. M., Dyke, J. P., Watts, R., Kobylarz, E. J., McCandliss, B. D., Heier, L. A., Beattie, B. J., Hamacher, K. A., Vallabhajosula, S., Goldsmith, S. J., Ballon, D., Giacino, J. T., & Schiff, N. D. (2006). Possible axonal regrowth in late recovery from the minimally conscious state. *J. Clin. Invest.*, *116*(7), 2005–2011.

Wang, J. W., David, D. J., Monckton, J. E., Battaglia, F., & Hen, R. (2008). Chronic fluoxetine stimulates maturation and synaptic plasticity of adult-born hippocampal granule cells. *J. Neurosci.*, *28*(6), 1374–1384.

Whyte, J., Katz, D., Long, D., DiPasquale, M. C., Polansky, M., Kalmar, K., Giacino, J., Childs, N., Mercer, W., Novak, P., Maurer, P., & Eifert, B. (2005). Predictors of outcome in prolonged posttraumatic disorders of consciousness and assessment of medication effects: A multicenter study. *Arch. Phys. Med. Rehabil.*, *86*(3), 453–462.

79 The Neurobiology of Consciousness

CHRISTOF KOCH

ABSTRACT Over the past two decades, the mystery of consciousness and its material basis has begun to be investigated by the sciences. The neurobiological approach to consciousness aims to identify its correlates at the neuronal level. Electrophysiological, psychophysical, and functional imaging studies in humans and nonhuman animals have allowed brain scientists to narrow their focus to the neural substrates of consciousness and conscious perception in the circuits of the forebrain, in particular the thalamocortical system and its satellites. These findings, complemented by the development of a robust theoretical predictive framework, should eventually lead to a rational understanding of the phenomenon of consciousness.

Consciousness is one of the most enigmatic features of the universe. People not only act but feel: they see, hear, smell, recall, plan for the future. These activities are associated with subjective, ineffable, immaterial feelings that are tied in some manner to the material brain. The exact nature of this relationship—the classical mind-body problem—remains elusive and the subject of heated debate (see Block, chapter 77, this volume). These firsthand, subjective experiences pose a daunting challenge to the scientific method that has, in many other areas, proven so immensely fruitful. The brute fact of consciousness comes as a total surprise; it does not appear to follow from any phenomena in traditional physics or biology.

People willingly concede that when it comes to nuclear physics or molecular biology, specialist knowledge is essential; but many assume that there are few relevant facts about consciousness, and therefore everybody is entitled to his or her own theory. Nothing could be further from the truth. There is an immense amount of relevant psychological, clinical, and neuroscientific data and observations that need to be accounted for.

Consciousness is a state-dependent property of certain complex, biological, adaptive, and highly interconnected systems. The best example of consciousness is found in a healthy and attentive human brain. In deep sleep, consciousness ceases. Small lesions in the midbrain, brain stem, and thalamus can lead to a complete loss of consciousness—probably through inactivation of the cerebral cortex (Sukhotinsky et al., 2007)—while destruction of circumscribed parts of the cerebral cortex can eliminate very specific aspects of consciousness, such as the ability to be aware of motion or to recognize faces, without a concomitant loss of vision in general. Brain scientists are focusing on experimental approaches that shed light on the neural basis of consciousness rather than on eristic philosophical arguments with no clear resolution. This chapter reviews these experimental approaches.

CHRISTOF KOCH Division of Biology, California Institute of Technology, Pasadena, California

What phenomena does consciousness encompass?

Consciousness has been dissected on conceptual grounds (access versus phenomenal consciousness; see Block, 2005, and chapter 77, this volume), ontological grounds (Hard versus Easy problem; Chalmers, 1996), and psychological grounds (explicit versus implicit processes; Tulving, 1993).

One common philosophical definition is "*Consciousness is what it is like to be something*," such as the experience of what it feels like to smell a rose or to be in love. This what-it-feels-like-from-within definition expresses the principal irreducible characteristic of the phenomenal aspect of consciousness: *to experience something*.

What it feels like to have a particular experience is called the *quale* of that experience: the quale of red is what is common to such disparate conscious states as seeing a red sunset, the red flag of China, arterial blood, or a ruby gemstone. All four subjective states share "redness." There are countless qualia (the plural of quale): the ways things look, sound, and smell, the way it feels to have a pain, the way it feels to have thoughts and desires, and so on. To have an experience means to have qualia, and the quale of an experience is what specifies it and makes it different from other experiences.

Science must explain the *exact relationship between the immaterial, conscious mind and its physical basis in the electrochemical interactions in the body*. This challenge can be decomposed into several subproblems.

Why is there any experience at all? Why does a brain state feel like anything? Many scholars have argued that the exact nature of this relationship will remain a central puzzle of

human existence, without an adequate reductionistic, scientific explanation. However, similar sentiments have been expressed in the past for the problem of seeking to understand life or to determine what material the stars are made of. Thus it is best to put this question aside for the moment and not be taken in by defeatist arguments.

Why is the relationship among different experiences the way it is? For instance, red, yellow, green, cyan, blue, and magenta are all colors that can be mapped onto the topology of a circle. Why? Furthermore, as a group, these color percepts share certain communalities that make them different from other percepts, such as seeing motion or smelling a rose. Why?

Why are feelings private? As expressed by poets and novelists, we cannot communicate an experience to somebody else except by way of example.

How do feelings acquire meaning? Subjective states are not abstract states but have an immense amount of associated explicit and implicit feelings. Think of the unmistakable smell of dogs coming in from the rain or the crunchy texture of potato chips. How do these arise?

Why are only some behaviors associated with conscious states? Much brain activity and many associated behaviors occur without any conscious sensation. Why? And where is the difference between the two at the neuronal level?

The neurobiology of free will

A further aspect of the mind-body problem is the question of free will and will power. Answering this question goes to the heart of the way people think of themselves. Of great relevance are the classical findings by Libet and colleagues (1983) of brain events that precede the conscious initiation of a voluntary action. This simple result has been replicated and extended but, because of its counterintuitive implication that conscious will has no causal role, continues to be vigorously debated (Haggard & Eimer, 1999; Soon, Brass, Heinze, & Haynes, 2008; Brass & Haggard, 2008; see Lau, chapter 83, this volume).

Psychological work in both normals and patients reveals dissociations between the conscious perception of a willed action and its actual execution: subjects believe that they perform actions that they did not do while, under different circumstances, subjects feel that they are not responsible for actions that are, demonstrably, their own (Wegner, 2002).

Whether volition is illusory or is free in some libertarian sense does not answer the question of how subjective states relate to brain states. The perception of free will, which psychologists call the *feeling of agency* or *authorship* (e.g., "I decided to lift my finger"), is a subjective state with an associated quale no different in kind from the quale of a toothache or seeing marine blue. It must also have some neuronal correlate. Direct electrical brain stimulation during neurosurgery, as well as fMRI experiments, implicates medial premotor and anterior cingulate cortices in generating the subjective feeling of triggering an action (see Lau, chapter 83, this volume).

Consciousness in other species

Data about conscious states come not only from people who can talk about their subjective experiences but also from nonlinguistic competent individuals—newborn babies (Lagerctantz & Changeux, 2009) or patients with complete paralysis of nearly all voluntary muscles (locked-in syndrome)—and, most importantly, from animals other than humans. There are at least three reasons to assume that many species, in particular those with complex behaviors such as mammals, share at least some aspects of consciousness with humans:

Similar neuronal architectures Except for size, there are no large-scale, dramatic differences between the cerebral cortex and thalamus of mice, monkeys, humans, and whales. It is difficult to distinguish a cubic millimeter of neocortex among different mammals, except by expert neuroanatomists.

Similar behavior Almost all human behaviors have precursors in the animal literature. Take the case of pain. The behaviors seen in humans when they experience pain and distress—facial contortions, moaning, yelping or other forms of vocalization, motor activity such as writhing, avoidance behaviors at the prospect of a repetition of the painful stimulus—can be observed in all mammals and in many other species. Likewise for the physiological signals that attend pain—activation of the sympathetic autonomous nervous system resulting in change in blood pressure, dilated pupils, sweating, increased heart rate, release of stress hormones, and so on. The discovery of cortical pain responses in premature babies shows the fallacy of relying on language as the sole criterion for consciousness (Slater et al., 2006).

Evolutionary continuity The first true mammals appeared at the end of the Triassic period, about 220 million years ago, with primates proliferating following the Cretaceous-Tertiary extinction event, about 60 million years ago, while humans and macaque monkeys did not diverge until 30 million years ago (Allman, 1999). *Homo sapiens* is part of an evolutionary continuum with its implied structural and behavioral continuity, rather than an independently developed organism.

While certain aspects of consciousness, in particular those relating to the recursive notion of self and to abstract, culturally transmitted knowledge, are not widespread in nonhuman animals, there is little reason to doubt that nonhuman mammals share conscious feelings—sentience—with humans. To believe that people are special, are singled out by the gift of consciousness above all other species, is a remnant of humanity's atavistic, deeply held belief that *Homo sapiens* occupies a privileged place in the universe, a belief with no empirical basis.

The extent to which nonmammalian vertebrates, such as tuna, cichlid, and other fish; crows, ravens, magpies, parrots, and other birds; or even invertebrates such as the octopus or bees, with complex, nonstereotyped behaviors including delayed matching, nonmatching to sample, and other forms of learning (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001) are conscious is difficult to answer at this point in time (Edelman, Baars, & Seth, 2005). Without a sounder understanding of the neuronal architecture necessary to support consciousness, it is unclear where in the animal kingdom to draw the Rubicon that separates species with at least some conscious percepts from those that never experience anything and that are nothing but pure automata (Griffin, 2001).

Arousal and states of consciousness

There are two common, but quite distinct, usages of the term consciousness, one revolving around *arousal* and *states of consciousness* (see Schiff, chapter 78, this volume) and another one around *the content of consciousness* and *conscious states*. To be conscious of anything, the brain must be in a relatively high state of arousal (sometimes also referred to as *vigilance*). This statement is as true of wakefulness as it is of REM sleep that is vividly, consciously experienced—though usually not remembered—in dreams. The level of brain arousal, measured by electrical or metabolic brain activity, fluctuates in a circadian manner, and is influenced by lack of sleep, drugs and alcohol, physical exertion, and so on in a predictable manner. High arousal states are always associated with some conscious state—a percept, thought, or memory—that has a specific *content*. We see a face, hear music, remember an incident, plan an experiment, or fantasize about sex. Indeed, it is not clear whether one can be awake without being conscious of something. Referring to such conscious states is conceptually quite distinct from referring to states of consciousness that fluctuate with different levels of arousal.

Different levels or states of consciousness are associated with different kinds of conscious experiences. The awake state in a normal functioning individual is quite different from the dreaming state (for instance, the latter has little or no self-reflection) or from the state of deep sleep. In all three cases, the basic physiology of the brain is changed, affecting the space of possible conscious experiences. Physiology is also different in *altered states of consciousness*, for instance, after taking psychedelic drugs when events often have a stronger emotional connotation than in normal life. Yet another state of consciousness can occur during certain meditative practices, when interoceptive perception and insight may be enhanced compared to the normal waking state.

In some obvious but difficult to rigorously define manner, the *richness of conscious experience* increases as an individual transitions from deep sleep to drowsiness to full wakefulness. This richness of possible conscious experience could be quantified using notions from complexity theory that incorporate both the dimensionality and the granularity of conscious experience (e.g., Tononi, 2004; see chapter 84, this volume). For example, inactivating all of visual cortex in an otherwise normal individual would significantly reduce the dimensionality of conscious experience, since no color, shape, motion, texture, or depth could be perceived or imagined. A singular exception to this progression is REM sleep where most motor activity is shut down in the *atonia* that is characteristic of this phase of sleep, and the person is difficult to wake up. Yet this low level of behavioral arousal goes, paradoxically, hand in hand with high metabolic and electrical brain activity and conscious, vivid states.

Clinicians speak of impaired states of consciousness as in "the comatose state," "the persistent vegetative state" (PVS), and the "minimal conscious state" (MCS). Here, state refers to different levels of consciousness, from a total absence in the case of coma, PVS, or general anesthesia, to a fluctuating and limited form of conscious sensation in MCS, in sleepwalking, or during a complex partial epileptic seizure (Schiff, 2004, and chapter 78, this volume)

The repertoire of distinct conscious states or experiences that are accessible to a patient in MCS is presumably minimal (possibly including pain, discomfort, and sporadic sensory percepts; but see Owen et al., 2006), immeasurably smaller than the possible conscious states that can be experienced by a healthy brain. Given the absence of any accepted theory for the minimal neuronal criteria necessary for consciousness, the distinction between a PVS patient—who shows regular sleep-wave transitions and who may be able to move eyes or limbs or smile in a reflexive manner, as in the widely publicized 2005 case of Terri Schiavo in Florida—and an MCS patient who can communicate (on occasion) in a meaningful manner (for instance, by differential eye movements) and who shows some signs of consciousness is often difficult to make in a clinical setting. Functional brain imaging of patients with global disturbances of consciousness (including akinetic mutism) reveal that dysfunction in a widespread cortical network including medial and lateral

prefrontal cortex and parietal associative areas is associated with a global loss of consciousness (Laureys, 2005).

In contrast to diffuse cortical damage, relatively discrete bilateral injuries to midline (paramedian) subcortical structures can also cause a complete loss of consciousness. These structures are therefore part of the *enabling* factors that control the level of brain arousal and that are needed for any form of consciousness to occur. For an example, consider the heterogeneous collection of more than two dozen (on each side) of nuclei in the upper brain stem (pons, midbrain, and posterior hypothalamus), collectively referred to as the *reticular activating system*. These nuclei—three-dimensional collections of neurons with their own cytoarchitecture and neurochemical identity—release distinct neuromodulators such as acetylcholine, noradrenaline/norepinephrine, serotonin, histamine, and orexin/hypocretin. Their axons project widely throughout the brain. These neuromodulators control the excitability of thalamus and forebrain and mediate the alternation between wakefulness and sleep, as well as the general level of both behavioral and brain arousal. Acute lesions in the reticular activating system can result in loss of consciousness and coma. Another enabling factor for consciousness is the *intralaminar nuclei of the thalamus* (ILN). These receive input from many brain stem nuclei and from frontal cortex and project strongly to the basal ganglia and, in a more distributed manner, into layer I of much of neocortex. Comparatively small (1 cm^3 or less) bilateral lesions in the ILN can completely eliminate awareness (Bogen, 1995). Thus, the ILN are necessary for somebody to be conscious at all but do not appear to be responsible for mediating specific conscious percepts. It is likely that the specific content of any one conscious sensation is mediated by neurons in cortex and their associated satellite structures, including the amygdala, thalamus, claustrum, and basal ganglia.

The neuronal correlates of consciousness

One key objective of the inchoate science of consciousness is to search for the neuronal correlates—and ultimately the causes—of consciousness. As defined by Crick and Koch (2003), the neuronal correlates of consciousness (NCC) are the *minimal neuronal mechanisms jointly sufficient for any one specific conscious percept* (figure 79.1).

This definition of the NCC stresses the word "minimal," because the question of interest is which subcomponents of the brain are actually needed. For instance, it is likely that neural activity in the cerebellum does not underlie any conscious perception and thus is not part of the NCC. That is, trains of spikes in Purkinje cells (or their absence) will not induce a sensory percept, although they may ultimately affect some behaviors.

This definition does not focus on the necessary conditions for consciousness, because of the great redundancy and parallelism found in neurobiological networks. While activity in some population of neurons may underpin a percept in one case, a different population might mediate a related percept if the former population is lost or inactivated.

Every phenomenal, subjective state will have associated NCC: one for seeing a red patch, another one for seeing Grandmother, yet a third one for feeling that a particular behavior was freely caused (feeling of agency). Perturbing or inactivating the NCC for any one specific conscious experience will affect the percept or cause it to disappear. If the NCC could be induced artificially—for instance, by cortical microstimulation in a prosthetic device or during neurosurgery—the subject will experience the associated percept. What characterizes the NCC? What are the communalities between the NCC for seeing and for hearing? Will the NCC involve all pyramidal neurons in cortex at any given point in time? Or only a varying subset of long-range projection

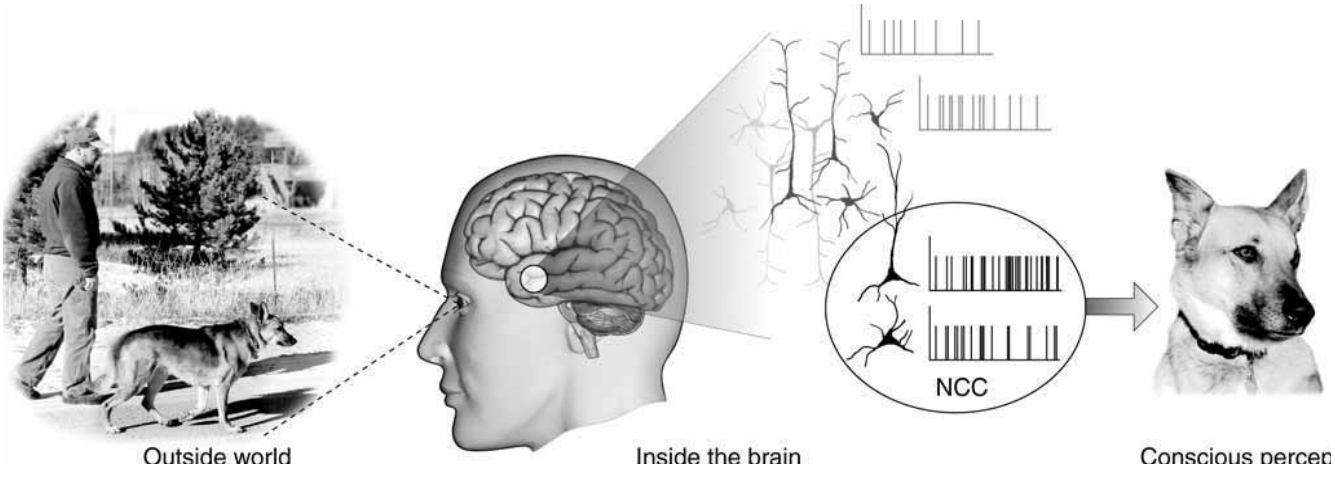


FIGURE 79.1 The neuronal correlates of consciousness (NCC) make up the minimal set of neural events and structures—here synchronized action potentials in neocortical pyramidal neurons—sufficient for a specific conscious percept or memory. (From Koch, 2004.)

cells in frontal lobes that project to the sensory cortices in the back? Only layer 5 cortical cells? Neurons that fire in a rhythmic manner? Neurons that fire in a synchronous manner? These are some of the proposals that have been advanced over the years (Chalmers, 2000).

The extent to which the NCC depends on emotions, moods, and homeostatic signals is controversial. This topic is taken up in detail by Koenigs and Adolphs, chapter 82, this volume).

Quantum mechanics and consciousness

It is implicitly assumed by neurobiologists that the relevant variables giving rise to consciousness are to be found at the neuronal level, among the synaptic releases or the action potentials in one or more population of cells, rather than at the molecular level. A few scholars have proposed that macroscopic quantum behaviors underlie consciousness. Of particular interest here is *entanglement*, the observation that the quantum states of multiple objects, such as two coupled electrons, may be highly correlated even though they are spatially separated, violating our intuition about locality (entanglement is also the key feature of quantum mechanics hoped to be exploited in quantum computers). The role of quantum mechanics for the photons received by the eye and for the molecules of life is not controversial. But there is no evidence that any components of the nervous system—a 37° Celsius wet and warm tissue strongly coupled to its environment—display quantum entanglement. And even if quantum entanglement were to occur inside individual cells, molecular diffusion and action potential generation and propagation, the principal mechanism for getting information into and out of neurons, would destroy superposition. At the cellular level, the interaction of neurons is governed by classical physics (Koch & Hepp, 2006).

Interaction with the world is not required for consciousness

We are usually conscious of what goes on around us, and occasionally of what goes on within our body. So it is only natural to think that consciousness may be tightly linked to the ongoing interaction we maintain with the world and the body (O'Regan & Noe, 2001). However, there are many examples to the contrary. We are conscious of our thoughts, which do not seem to correspond to anything out there; we can also imagine things that are not out there. When we do so, cortical sensory areas can be activated from the inside, though there are some differences from normal visual perception. Also, stimulus-independent consciousness is associated with its own patterns of activation within cortex and thalamus (Mason, Norton, Van Horn, Wegner, & Grafton, 2007). During dreams, we are virtually disconnected from the environment (Hobson, Pace-Schott, & Stickgold, 2000)—hardly anything of what happens around us enters consciousness, and our muscles are paralyzed (except for eye muscles and diaphragm). Nevertheless, we are vividly conscious: all that seems to matter is that the thalamocortical system continues to function more or less as in wakefulness, as shown by neuronal recording, EEG, and neuroimaging studies performed during rapid eye movement (REM) sleep, when dreams are most intense (Maquet et al., 1996).

Neurological evidence indicates that neither sensory inputs nor motor outputs are needed to generate consciousness. For instance, retinally blind people can both imagine and dream visually if they become blind after 6–7 years of age or so (Hollins, 1985; Buchel, Price, Frackowiak, & Friston, 1998). Patients with the locked-in syndrome can be almost completely paralyzed, and yet they are just as conscious as healthy subjects (Laureys, 2005) and can compose eloquent accounts of their condition (Bauby, 1997). A transient form of paralysis is one of the characteristic features of narcolepsy. Severe cataleptic attacks can last for minutes and leave the patient collapsed on the floor, utterly unable to move or to signal, but fully aware of her surroundings (Siegel, 2000). Or consider the Californian drug addicts known as the *frozen addicts* who acquired some of the symptoms of severe, late-stage Parkinson's disease, fully conscious, yet unable to move or speak (Langston & Palfreman, 1995). All six had previously taken synthetic heroin tainted with MPTP, which selectively and permanently destroyed dopamine-producing neurons in their basal ganglia.

Consciousness here and now depends on what certain parts of the brain are doing, without requiring any obligatory interaction with the environment or the body. Whether the *development* of consciousness requires such interactions in early childhood, though, is a different matter.

Consciousness does not require self-consciousness, introspection, or language

Consciousness is usually evaluated by verbal reports. Questions about consciousness ("Did you see anything on the screen?") are answered by "looking inside" retrospectively and reporting what one has just experienced. So it is perhaps natural to suggest that consciousness may arise through the ability to reflect on our own perceptions: our brain would form a scene of what it sees, but we would become conscious of it—experience it subjectively—only when we, as a subject of experience, watch that scene from the inside. This suggestion is often framed in a neurobiological context by assuming that patterns of activity corresponding to "unconscious" or "subconscious" percepts form in posterior regions of the cerebral cortex involved in the categorization/association of sensory stimuli. These percepts then become conscious when mainly

anterior prefrontal and cingulate regions involved in self-representations and introspection interact with posterior cortex, perhaps by reading signals through forward connections and selectively amplifying them through back connections.

There is no doubt that the brain categorizes its own patterns of activity in the sense that neurons respond mainly to the activity of other neurons, so the brain is constantly "looking at itself." However, this process is not necessarily understood in terms of a "subject" (the front) looking at an "object" represented in sensory cortices (the back). Leaving aside the mystery of why reflecting on something should make it conscious, this scenario is made less plausible by a common observation: when we become absorbed in some engaging task—for example, watching an engrossing movie, playing a fast-paced video game, or driving a motorcycle at high speed through traffic—we are vividly conscious without reflection or introspection. Often, we become so immersed in this rapid flow of experience—for example, during a difficult climb up a rock wall—that we may lose the sense of self, the inner voice. Perhaps the habit of thinking about consciousness has distracted the scholars who write upon such matters to devalue the unreflective nature of much of experience.

A neuroimaging study by Malach and collaborators (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004) suggests that activation of prefrontal regions is not necessary for the emergence of perceptual consciousness but may be needed to reflect upon it and report it to others (however, see Bar et al., 2006 for evidence of very rapid and visual object-specific activation of orbitofrontal cortex). Indeed, it appears that self-related activity is actually shut off during highly demanding sensory tasks. Lesion studies also support the notion that perceptual consciousness may not require prefrontal cortex and, by inference, the functions it performs: A man who, at the age of 21, had fallen on an iron spike that completely penetrated through both of his frontal lobes, nevertheless went on to live a stable life—marrying and raising two children—in an appropriate professional and social setting. Although displaying many of the typical frontal lobe behavioral disturbances, he never complained of loss of sensory perception, nor did he show visual or other deficits (Mataró, Jurado, García-Sánchez, Barraquer, & Costa-Jussa, 2001). Another case is that of a 27-year-old woman with massive bilateral prefrontal damage of unclear etiology (Markowitsch & Kessler, 2000). While manifesting grossly deficient scores in frontal-lobe-sensitive tests, she has no abnormal perceptual abilities (that is not to say that such patients do not suffer from subtle visual deficits; Barcelo, Suwazono, & Knight, 2000).

Finally, being conscious does not require language. Humans continually affirm consciousness through speech, describing and discussing their sensory and other experiences. So it is natural to think that speech and consciousness are somehow inextricably linked. They are not. Infants and animals cannot speak, but they are conscious and can report their experiences in other ways. And, of course, there are numerous patients who lost the ability to understand or use words and yet remained conscious.

Consciousness and attention are independent processes

Few would dispute that the relationship between consciousness and selective, visual attention is an intimate one. When subjects pay attention to an object, they become conscious of its various attributes; when the focus of attention shifts away, the object fades from consciousness. Indeed, more than a century of research efforts have quantified the ample benefits accrued to attended and consciously perceived events (Pashler, 1998; Braun, Koch, & Davis, 2001). This intimate connection has prompted many to posit that the two processes are inextricably interwoven, if not identical (Posner, 1994; Merikle & Joordens, 1997; Chun & Wolfe, 2000; O'Regan & Noe, 2001). Others, however, going back to the 19th century, have argued that attention and consciousness are distinct phenomena, with distinct functions and neuronal mechanisms (Iwasaki, 1993; Hardcastle, 1997; Lamme, 2003; Baars, 2005; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Koch & Tsuchiya, 2007). Recent psychophysical and neurophysiological evidence argues in favor of a dissociation between selective attention and consciousness, suggesting that events or objects can be attended to without being consciously perceived. Conversely, an event or object can be consciously perceived in the near absence of top-down attentional processing.

Attention Without Consciousness Consider that subjects can attend to a location for many seconds and yet fail to see one or more attributes of an object at that location. In lateral masking (visual crowding), the orientation of a peripherally presented grating is hidden from conscious sight but remains sufficiently potent to induce an orientation-dependent aftereffect (He, Cavanagh, & Intriligator, 1996). Montaser-Kouhsari and Rajimehr (2004) showed that an aftereffect induced by an invisible illusory contour required focal attention, even though the object at the center of attention was invisible. Naccache, Blandin, and Dehaene (2002) elicited priming for invisible words (suppressed by forward and backward masking) but only if the subject was attending to the invisible prime; without attention, the same word failed to elicit priming. In another experiment, male/female nudes attracted attention when they were rendered completely invisible by continuous flash suppression (Jiang, Costello, Fang, Huang, & He, 2006). When subjects had to discriminate the location of the masked nude from the location of a masked shuffled nude, they were at chance; without the intraocular masking, the images are clearly

visible. Functional MRI evidence confirms attentional modulation of invisible images in primary visual cortex (Bahrami, Lavie, & Rees, 2007). In conclusion, attentional selection by itself is not sufficient for consciousness to occur.

Consciousness in the Absence of Attention When one focuses intensely on one event, the world is not reduced to a tunnel, with everything outside the focus of attention gone: we are always aware of some aspects of the world surrounding us, such as its gist. Indeed, gist is immune from inattentional blindness (Mack & Rock, 1998). In the 30 ms necessary to apprehend the gist of a scene, top-down attention cannot play much of a role (because gist is a property associated with the entire image, any process that locally enhances features is going to be of limited use; Fei-Fei, Iver, Koch, & Perona, 2007).

Take the perception of a single object (say a bar) in an otherwise empty display, a nonecological but common arrangement in many animal and human experiments. Here, what function would top-down, selective attention need to perform without any competing objects nearby? Indeed, the most popular neuronal model of attention, biased competition (Desimone & Duncan, 1995), predicts that in the absence of competition no or little attentional enhancement occurs, yet we are perfectly aware of the object and its background.

In a dual-task paradigm, the subject's attention is drawn to a demanding central task, while at the same time a secondary stimulus is flashed somewhere in the periphery. Using the identical retinal layout, the subject performs either the central task, or the peripheral task, or both simultaneously (Sperling & Dosher, 1986; Braun & Sagi, 1990; Braun & Julesz, 1998). With focal attention engaged at the center, the subject can still distinguish a natural scene containing an animal (or a vehicle) from one that does not include an animal (or a vehicle), while being unable to discriminate a red-green bisected disk from a green-red one (Li et al., 2002). Likewise, subjects can tell male from female faces or even distinguish a famous from a nonfamous face (Reddy, Wilken, & Koch, 2004; Reddy, Reddy, & Koch, 2006), but are frustrated by computationally much simpler tasks (e.g., discriminating a rotated letter *L* from a rotated *T*). Thus, although we cannot be sure that observers do not deploy some limited amount of top-down attention in dual-task experiments that require training and concentration (that is, high arousal), it remains true that subjects can perform certain discriminations but not others in the near absence of top-down attention. And they are not guessing. They can be quite confident of their choices and "see," albeit often indistinctly, what they can discriminate.

The existence of such dissociations—attention without consciousness and consciousness without attention—should not be surprising when considering their different functions. Attention is the set of mechanisms whereby the brain selects a subset of the incoming sensory information for higher-level processing, while the nonattended portion of the input is analyzed at a lower bandwidth, that is, with fewer processing resources. In primates, about one million fibers leave each eye and carry on the order of one megabyte per second of raw information. One way to deal with this deluge of data is to select a small fraction and process this reduced input in real time while nonattended stimuli suffer from benign neglect. By contrast, consciousness appears to be involved in providing a kind of "executive summary" of the current situation that is useful for decision making, planning, and learning (Koch, 2004; Baars, 2005).

The neuronal basis of perceptual illusions

The possibility of precisely manipulating visual percepts in time and space has made vision a preferred modality for seeking the NCC. Psychologists have perfected a number of techniques—masking, binocular rivalry, continuous flash suppression, motion-induced blindness, change blindness, inattentional blindness—in which the seemingly simple and unambiguous relationship between a physical stimulus in the world and its associated percept in the privacy of the subject's mind is disrupted (Kim & Blake, 2005; see also Rees, chapter 80 in this volume, and Macknik & Martinez-Conde, chapter 81). A stimulus can be perceptually suppressed for minutes at a time: the image is projected into one of the observer's eyes, but it is invisible, not seen. In this manner the neural mechanisms that respond to the subjective percept rather than the physical stimulus can be isolated, permitting the footprints of visual consciousness to be tracked in the brain.

A popular illusion is binocular rivalry (Blake & Logothetis, 2002). Here, a small image (e.g., a horizontal grating) is presented to the left eye and another image (e.g., a vertical grating) is shown to the corresponding location in the right eye. In spite of the constant visual stimulus, observers consciously see the horizontal grating alternate every few seconds with the vertical one. The brain does not allow for the simultaneous perception of both images.

Macaque monkeys can be trained to report whether they see one or the other image. The distribution of switching times and the way in which changing the contrast of one image affects the reports leave little doubt that monkeys and humans experience the same basic phenomenon. In a series of elegant experiments, Logothetis and colleagues (Leopold & Logothetis, 1996; Logothetis, 1998) recorded from a variety of visual cortical areas in the awake macaque monkey while the animal performed a binocular rivalry task. In primary visual cortex (V1), only a small fraction of cells weakly modulate their response as a function of the percept of the monkey. The majority of cells responded to one or the

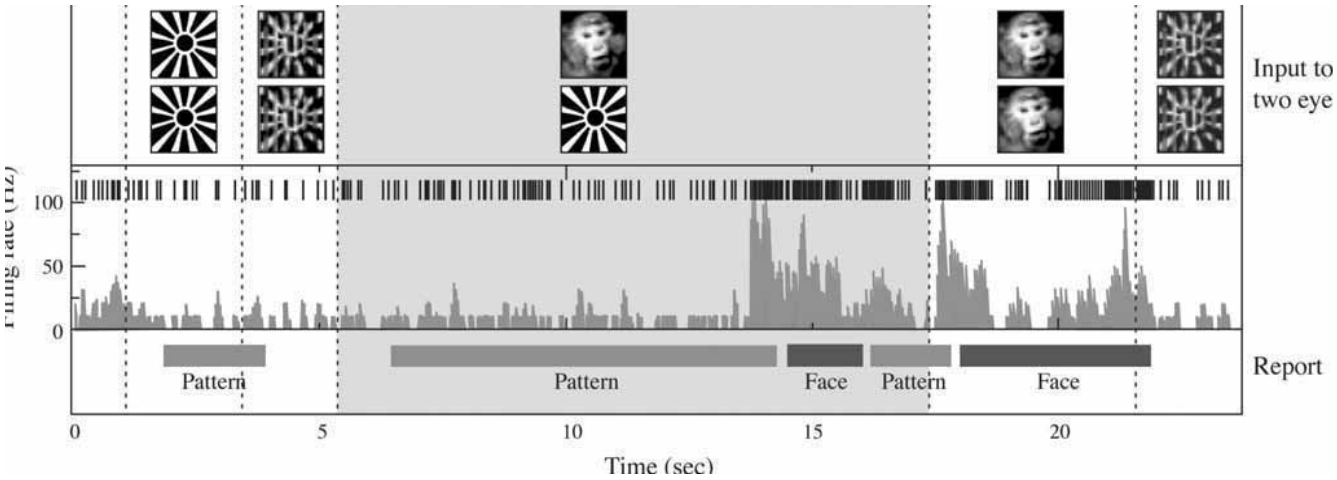


FIGURE 79.2 A fraction of a minute in the life of a typical IT cell while a monkey experiences binocular rivalry. The upper row indicates the visual input, with dotted vertical boundaries marking stimulus transitions. The second row shows the individual spikes, the third row the smoothed firing rate, and the bottom row the monkey's behavior. The animal was taught to press a lever when it saw either one or the other image, but not both. The cell responded only weakly to either the sunburst design or to its optical superposition with the image of a monkey's face. During binocular rivalry (gray zone), the monkey's perception vacillated back and forth between seeing the face and seeing the bursting sun. Perception of the face was consistently accompanied (and preceded) by a strong increase in firing rate. (From N. Logothetis, private communication, as modified by Koch, 2004.)

other retinal stimulus with little regard to what the animal perceived at the time. In contrast, in a high-level cortical area such as the inferior temporal (IT) cortex along the ventral pathway, almost all neurons responded only to the perceptual dominant stimulus, that is, to the stimulus that was being reported. For example, when a face and a more abstract design were presented, one of these to each eye, a "face" cell fired only when the animal indicated by its performance that it saw the face and not the design presented to the other eye (figure 79.2). This result implies that the NCC involve activity in neurons in inferior temporal cortex.

Clearly this does not imply that the NCC are local to IT. Given known anatomical connections, it is likely that specific reciprocal interactions between IT cells and neurons in parts of the prefrontal cortex are necessary for the NCC. This possibility is compatible with the widely accepted notion that the NCC involve positive feedback to ensure that neural activity is persistent and strong enough to exceed some threshold and to be distributed to multiple cognitive systems, including working memory, planning, and language.

In a related perceptual phenomenon, *flash suppression*, the percept associated with an image projected into one eye is transiently suppressed by flashing another image into the other eye (while the original image remains; Wolfe, 1984). Its methodological advantage over binocular rivalry is that the timing of the perceptual transition is determined by an external trigger rather than by an internal event. The majority of responsive cells in inferior temporal cortex and in the superior temporal sulcus follow the monkey's behavior—and therefore its percept (Sheinberg & Logothetis, 1997). That is, when the animal perceives a cell's preferred stimulus, the neuron fires; when the stimulus is present on the retina but is perceptually suppressed, the cell falls silent, even though legions of V1 neurons fire vigorously to the same stimulus. Single-neuron recordings in the medial temporal lobe of epileptic patients during flash suppression likewise demonstrate abolition of their responses when their preferred stimulus is present on the retina but not seen (Kreiman, Fried, & Koch, 2002).

A related question is the extent to which a specialized network of neurons in any one cortical region mediates the NCC for all columnar properties associated with that region. This has been directly tested by recording from individual neurons in the middle temporal cortex (MT) of monkeys viewing perceptual rivalrous motion stimuli (Maier, Logothetis, & Leopold, 2007). Contrary to expectations, small changes in the stimulus configuration lead to large changes in the firing activity of cells that carry perceptual rather that purely sensory signals. Depending on which one of four stimulus configurations the physiologists used, between 70% and 90% of all MT cells can carry NCC-related signals. This result implies either that specialized cells expressing the NCC are located beyond area MT or that such specialized cells do not exist in large numbers and that almost any neuron can participate in mediating perceptual consciousness.

A number of fMRI experiments have exploited binocular rivalry and related illusions to identify the hemodynamic activity underlying visual consciousness in humans. They demonstrate quite conclusively that BOLD activity in the upper stages of the ventral pathway (e.g., the fusiform face area and the parahippocampal place area) follow the percept

and not simply the retinal stimulus (Rees & Frith, 2007; Rees, chapter 80, this volume).

There is a lively debate about the extent to which neurons in primary visual cortex are directly responsible for expressing the subject's conscious percept. That is, is V1 part of the NCC (Crick & Koch, 1995)? It is clear that retinal neurons are not part of the NCC for visual experiences. While retinal neurons often correlate with visual experience, the spiking activity of retinal ganglion cells does not accord with visual experience (for example, there are no photoreceptors at the blind spot; yet no hole in the field of view is apparent; in dreams, vivid imagery occurs despite closed eyes; and so on).

A number of compelling observations link perception with fMRI BOLD activity in human V1 and even LGN (Tong, Nakayama, Vaughan, & Kanwisher, 1998; Lee, Blake, & Heeger, 2005). These data appear to be at odds with single-neuron recordings from the monkey (but see Maier et al., 2008). It is known that modulatory, feedback signals—for example those mediating selective attention—can be much more easily detected by means of fMRI than by single-unit recordings (Wilke, Logothetis, & Leopold, 2006; Logothetis, 2008). Indeed, unless attentional effects are carefully controlled for, their neural correlates cannot be untangled from those of consciousness (Huk, Rees, & Heeger, 2001; Tse, Martinez-Conde, Schlegel, & Macknik, 2005). This aim has now been achieved in an elegant study of Lee, Blake, and Heeger (2007). Using a dual-task paradigm they find that hemodynamic BOLD activity in human V1 reflects attentional processes but does not directly correlate with the conscious percept of the subject. Haynes and Rees (2005) exploited multivariate decoding techniques to read out perceptually suppressed information (the orientation of a masked stimulus) from V1 BOLD activity, even though the stimulus orientation was so efficiently masked that subjects performed at chance levels when guessing the orientation. This finding supports the hypothesis that information present in V1 is accessible neither to behavior nor to consciousness, as hypothesized by Crick and Koch (1995).

In a powerful combination of binocular rivalry and flash suppression, a stationary image in one eye can be suppressed for minutes on end by continuously flashing different images into the other eye (*continuous flash suppression*; Tsuchiya & Koch, 2005; Tsuchiya, Koch, Gilroy, & Blake, 2006). This paradigm lends itself naturally to further investigate the relationship between neural activity—whether assayed at the single-neuron or at the brain-voxel level—and conscious perception (Jiang & He, 2006).

Other questions related to perceptual consciousness

The attributes of even simple percepts seem to vary along a continuum. For instance, a patch of color has a brightness and a hue that are variable, just as a simple tone has an associated loudness and pitch. However, is it possible that each particular, consciously experienced, percept is all-or-none? Might a pure tone of a particular pitch and loudness be experienced as an atom of perception, either heard or not, rather than gradually emerging from the noisy background? The perception of the world around us would then be a superposition of many elementary, binary percepts (Sergent & Dehaene, 2004).

Is perception continuous, like a river, or does it consist of a series of discontinuous batches, rather like the discrete frames in a movie (Purves, Paydarfar, & Andrews, 1996; VanRullen & Koch, 2003)? In *cinematographic vision* (Sacks, 2004), a rare form of visual migraine, the subject sees the movement of objects as fractured in time, as a succession of different configurations and positions, without any movement in between. The hypothesis that visual perception is quantized in discrete batches of variable duration, most often related to EEG rhythms in various frequency ranges (from theta to beta), is an old one. This idea is being revisited in light of the discrepancies of timing of perceptual events within and across different sensory modalities. For instance, even though a change in the color of an object occurs simultaneously with a change in its direction of motion, it may not be perceived that way (Zeki, 1998; Bartels & Zeki, 2006; Stetson, Cui, Montague, & Eagleman, 2006).

Forward versus feedback projections

Many actions in response to sensory inputs are rapid, transient, stereotyped, and unconscious (Milner & Goodale, 1995). They can be thought of as cortical reflexes and are sometimes called *zombie behaviors* (Koch & Crick, 2001). A slower, all-purpose conscious mode deals with broader, less stereotyped, and more complex aspects of the sensory input (or a reflection of these, as in imagery) and takes time to decide on appropriate responses. A consciousness mode is needed because otherwise a vast number of different zombie modes would be required to react to unusual events. The conscious system may interfere somewhat with the concurrent zombie systems (Beilock, Carr, MacMahon, & Starkes, 2002): focusing consciousness onto the smooth execution of a complex, multi-component, and highly trained sensorimotor task—dribbling a soccer ball, to give one example—can interfere with its smooth execution, something well known to athletes and their trainers. Having both a zombie mode that responds in a well-rehearsed and stereotyped manner and a slower system that allows time for planning more complex behavior is a great evolutionary discovery. This latter aspect, planning, may be one of the principal functions of consciousness.

It seems possible that visual zombie modes in the cortex mainly use the dorsal stream in the parietal region (Milner

& Goodale, 1995). However, parietal activity can affect consciousness by producing attentional effects on the ventral stream, at least under some circumstances. The basis of this inference is clinical case studies and fMRI experiments in normal subjects (Corbetta & Shulman, 2002). The conscious mode for vision depends largely on the ventral "what" stream (but see Bar et al., 2006).

Seemingly complex visual processing (such as detecting animals in natural, cluttered images) can be accomplished by cortex within 130–150 ms (Thorpe, Fize, & Marlot, 1996; VanRullen & Koch, 2003), too fast for consciousness to occur. It is plausible that such behaviors are mediated by a purely feedforward moving wave of spiking activity that passes from the retina through V1, into V4, IT, and prefrontal cortex, until it affects motor neurons in the spinal cord that control the finger press (as in a typical laboratory experiment). The hypothesis that the basic processing of information is feedforward is supported most directly by the short times required for a selective response to appear in IT cells (Perrett, Hietanen, Oram, & Benson, 1992). Indeed, Hung and colleagues (2005) were able to decode from the spiking activity 100 ms after image onset from a couple of hundred neurons in monkey IT the identity of a single image flashed onto the retina of the fixating animal. Coupled with a suitable motor output, such a feedforward network implements a zombie behavior—rapidly and efficiently subserving a binary categorization task in the absence of any conscious experience.

Conscious perception is believed to require more sustained, reverberatory neural activity, most likely by way of cortico-cortical feedback from other neocortical regions (see Macknik & Martinez-Conde, chapter 81, this volume). These feedback loops would explain why in backward masking a second stimulus, flashed 80–100 ms after onset of a first image, can still interfere (mask) with the percept of the first image. The reverberatory activity builds up over time until it exceeds a critical threshold. At this point, the sustained neural activity rapidly propagates to parietal, prefrontal, and anterior cingulate cortical regions, thalamus, claustrum (Crick & Koch, 2005), and related structures that support short-term memory, multimodality integration, planning, speech, and other processes intimately related to consciousness. Competition prevents more than one or a very small number of percepts to be simultaneously and actively represented. This is the hypothesis at the heart of the *global workspace* model of consciousness (Baars, 1988; Dehaene, Sergent, & Changeux, 2003). Sending visual information to more frontal structures would allow the associated visual events to be decoded and placed into context (for instance, by accessing various memory banks) and to have this interpretation feed back to the stimulus representation in visual cortex (Jazayeri & Movshon, 2007).

Conclusion

Ever since the Greeks first considered the mind-body problem more than two millennia ago, it has been the domain of armchair speculations and esoteric debates with no apparent resolution. Yet many aspects of this ancient set of questions now fall squarely within the domain of science.

It is known that consciousness does not require sensory input or motor output. Based on clinical and brain-imaging evidence, consciousness does not require self-consciousness, reflection, introspection, or language, although all these capabilities deeply enrich consciousness. Psychophysical and imaging evidence demonstrates that consciousness and selective attention can be dissociated. It appears that the neuronal correlates of consciousness require extensive but selective activity in the thalamocortical system, supported by enabling systems in the central thalamus, midbrain, and brain stem.

To further progress, it is imperative to record from a large number of neurons simultaneously at many locations throughout the thalamocortical system and related satellites (in particular the claustrum; Crick & Koch, 2005) in behaving subjects. This effort also demands a battery of behaviors (akin to but different from the well-known Turing test for intelligence; Koch & Tononi, 2008) that the subject—a newborn infant, immobilized patient, or nonhuman animal—has to pass before considering him, her, or it to possess some measure of consciousness. This is not an insurmountable step for mammals such as the monkey or the mouse that share many behaviors and brain structures with humans. For example, one particular mouse model of contingency awareness (C. Han et al., 2003) is based on the differential requirement for awareness of trace versus delay associative eyeblink conditioning in humans (Clark & Squire, 1998).

The growing ability of neuroscientists to manipulate in a reversible, transient, deliberate, and delicate manner identified populations of neurons using methods from molecular biology (Aravanis, Wang, Zhang, Meltzer, & Mogri, 2007; X. Han & Boyden, 2007; Zhang, Wang, Adamantidis, de Lecea, & Deisseroth, 2007) opens the possibility of moving from correlation—observing that a particular conscious state is associated with some neural or hemodynamic activity—to causation. Exploiting these increasingly powerful tools depends on the simultaneous development of appropriate behavioral assays and model organisms amenable to large-scale genomic analysis and manipulation, particularly in mice (Lein et al., 2007).

Finally, as mentioned previously, it is not known to what extent animals whose nervous systems have an architecture considerably different from the mammalian neocortex are conscious. Furthermore, whether artificial systems, such as

computers, robots, or the World Wide Web as a whole, which behave with considerable intelligence, are or can become conscious remains speculative (Koch & Tononi, 2008). What is needed is a theory of consciousness that explains in quantitative terms what type of systems, with what architecture, can possess conscious states.

Information theory may be such a theoretical approach that establishes at the fundamental level what consciousness is, how it can be measured, and what requisites a physical system must satisfy in order to generate it (Chalmers, 1996; Tononi & Edelman, 1998).

The most promising candidate for such a theoretical framework is the *integrated information theory of* consciousness discussed in more detail by Tononi and Balduzzi (chapter 84, this volume).

It is the combination of fine-grained neuronal analysis in animals, with ever more sensitive psychophysical and brain-imaging techniques in patients and healthy individuals, and the development of a robust theoretical framework that lend hope that we can ultimately understand one of the central mysteries of life.

REFERENCES

Allman, J. M. (1999). *Evolving brains*. New York: Scientific American.

Aravanis, A. M., Wang, L. P., Zhang, F., Meltzer, L. A., & Mogri, M. Z. (2007). An optical neural interface: In vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J. Neural Eng.*, *4*(3), S143–156.

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Prog. Brain Res.*, *150*, 45–53.

Bahrami, B., Lavie, N., & Rees, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Curr. Biol.*, *17*(6), 509–513.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). "Top-down facilitation of visual recognition." *Proc. Natl. Acad. Sci. USA 103*(2), 449–454.

Barcelo, F., Suwazono, S., & Knight. R. T. (2000). Prefrontal modulation of visual processing in humans. *Nat. Neurosci.*, *3*(4), 399–403.

Bartels, A., & Zeki, S. (2006). "The temporal order of blinding visual attributes." *Vision Research 46*, 2280–2286.

Bauby, J.-D. (1997). *The diving-bell and the butterfly: A memoir of life in death*. New York: Alfred A. Knopf.

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of dividend versus skill-focused attention on novice and experienced performance of sensorimotor skills. *J. Exp. Psychol. Appl.*, *8*, 6–16.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nat. Rev. Neurosci.*, *3*, 13–21.

Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci.*, *9*, 46–52.

Bogen, J. E. (1995). On the neurophysiology of consciousness. I. An overview. *Conscious. Cogn.*, *4*, 52–62.

Brass, M., & Haggard, P. (2008). To do or not to do: The neural signature of self-control. *J. Neurosci.*, *27*, 9141–9145.

Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: Detection and discrimination tasks. *Percept. Psychophys.*, *60*(1), 1–23.

Braun, J., Koch, C., & Davis, J. (2001). *Visual attention and cortical circuits*. Cambridge, MA: MIT Press.

Braun, J., & Sagi, D. (1990). Vision outside the focus of attention. *Percept. Psychophys.*, *48*(1), 45–58.

Buchel, C., Price, C., Frackowiak, R. S., & Friston, K. (1998). Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain*, *121*(Pt. 3), 409–419.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press.

Chalmers, D. J. (2000). *What is a neural correlate of consciousness?* Cambridge, MA: MIT Press.

Chun, M. M., & Wolfe, J. M. (2000). Visual attention. In E. B. Goldstein (Ed.), *Blackwell handbook of perception* (pp. 272–310). Hoboken, NJ: Wiley-Blackwell.

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science*, *280*(5360), 77–81.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.*, *3*, 201–215.

Crick, F. C., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, *375*, 121–123.

Crick, F. C., & Koch, C. (2003). A framework for consciousness. *Nat. Neurosci.*, *6*, 119–127.

Crick, F. C., & Koch, C. (2005). What is the function of the claustrum. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *360*, 1271–1279.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends Cogn. Sci.*, *10*(5), 204–211.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. USA*, *100*(14): 8520–8525.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, *18*, 193–222.

Edelman, D. B., Baars, J. B., & Seth, A. K. (2005). Identifying hallmarks of consciousness in non-mammalian species. *Conscious. Cogn.*, *14*, 169–187.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *J. Vis.*, *7*(10).

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of "sameness" and "difference" in an insect. *Nature*, *410*, 930–933.

Griffin, D. R. (2001). *Animal minds: Beyond cognition to consciousness*. Chicago: University of Chicago Press.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and conscious awareness. *Exp. Brain Res.*, *126*, 128–133.

Han, C. J., O'Tuathaigh, C. M., van Trigt, L., Quinn, J. J., Fanselow, M. S., Mongeau, R., Koch, C., & Anderson, D. J. (2003). Trace but not delay fear conditioning requires attention

and the anterior cingulate cortex. *Proc. Natl. Acad. Sci. USA*, *100*, 13087–13092.

Han, X., & Boyden, E. S. (2007). Multiple-color optical activation, silencing, and desynchronization of neural activity, with single-spike temporal resolution. *PLoS One*, *2*(3), e299.

Hardcastle, V. G. (1997). Attention versus consciousness: A distinction with a difference. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, *4*, 56–66.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640.

Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.*, *8*(5), 686–691.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*(6598), 334–337.

Hobson, J. A., Pace-Schott, E. F., & Stickgold, R. (2000). Dreaming and the brain: Toward a cognitive neuroscience of conscious states. *Behav. Brain Sci.*, *23*(6), 793–842; discussion, 904–1121.

Hollins, M. (1985). Styles of mental imagery in blind adults. *Neuropsychologia*, *23*(4), 561–566.

Huk, A. C., Ress, D., & Heeger, D. (2001). Neuronal basis of the motion aftereffect reconsidered. *Neuron*, *32*(1), 161–172.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, *310*(5749), 863–866.

Iwasaki, S. (1993). Spatial attention and two modes of visual consciousness. *Cognition*, *49*(3), 211–233.

Jazayeri, A., & Movshon, J. A. (2007). A new preceptual illusion reveals mechanisms of sensory decoding. *Nature*, *446*, 912–915.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proc. Natl. Acad. Sci. USA*, *103*, 17048–17052.

Jiang, Y., & He, S. (2006). Cortical responses to invisible faces: Dissociating subsystems for facial-information processing. *Curr. Biol.*, *16*, 2023–2029.

Kim, C.-Y., & Blake, R. (2005). Psychophysical magic: Rendering the visible "invisible." *Trends Cogn. Sci.*, *9*, 381–388.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Denver, CO: Roberts.

Koch, C., & Crick, F. C. (2001). On the zombie within. *Nature*, *411*, 893.

Koch, C., & Hepp, K. (2006). Quantum mechanics and higher brain functions: Lessons from quantum computation and neurobiology. *Nature*, *440*, 611–612.

Koch, C., & Tononi, G. (2008). Can machines be conscious? *IEEE Spectrum*, *45*, 54–59.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends Cogn. Sci.*, *11*, 16–22.

Kreiman, G., Fried, I., & Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc. Natl. Acad. Sci. USA*, *99*, 8378–8383.

Lagercrantz, H., & Changeux, J.-P. (2009). The emergence of human consciousness: From fetal to neonatal life. *Pediatr. Res.*, *65*, 255–260.

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci.*, *7*(1), 12–18.

Langston, J., & Palfreman, J. (1995). *The case of the frozen addicts*. New York: Vintage Books.

Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends Cogn. Sci.*, *9*, 556–559.

Lee, S. H., Blake, R., & Heeger, D. (2005). Traveling waves of activity in primary visual cortex. *Nat. Neurosci.*, *8*, 22–23.

Lee, S. H., Blake, R., & Heeger, D. (2007). Hierarchy of cortical responses underlying binocular rivalry. *Nat. Neurosci.*, *10*(8), 1048–1054.

Lein, E. S., Hawrylycz, M. J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, *445*, 168–176.

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflects monkeys' percepts during binocular rivalry. *Nature*, *379*, 549–553.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. USA*, *99*(14), 9596–9601.

Libet, B., Gleason, C. A.,Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*, *106*, 623–642.

Logothetis, N. K. (1998). Single units and conscious vision. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *353*, 1801–1818.

Logothetis, N. K. (2008). What can we do and what we cannot do with fMRI. *Nature*, *453*, 869–878.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Maier, A., Logothetis, N. K., & Leopold, D. A. (2007). Context-dependent perceptual modulation of single neurons in primate visual cortex. *Proc. Natl. Acad. Sci. USA*, *104*, 5620–5625.

Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F. Q., & Leopold, D. A. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat. Neurosci.*, *11*, 1193–1200.

Maquet, P., Péters, J., Peters, J.-M., Aerts, J., Delfiore, G., Degueldre, C., Luxen, A., & Franck, G. (1996). Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. *Nature*, *383*(6596), 163–166.

Markowitsch, H. J., & Kessler, J. (2000). Massive impairment in executive functions with partial preservation of other cognitive functions: The case of a young patient with severe degeneration of the prefrontal cortex. *Exp. Brain Res.*, *133*(1), 94–102.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., & Grafton, S. T. (2007). Wandering minds: The default network and stimulus-independent thought. *Science*, *315*(5810), 393–395.

Mataró, M., Jurado, M. A., García-Sánchez, C., Barraquer, L., Costa-Jussa, F. R., & Junqué, C. (2001). Long-term effects of bilateral frontal brain lesion: 60 years after injury with an iron bar. *Arch. Neurol.*, *58*(7), 1139–1142.

Merikle, P. M., & Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Conscious. Cogn.*, *6*(2–3), 219–236.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford, UK: Oxford University Press.

Montaser-Kouhsari, L., & Rajimehr, R. (2004). Attentional modulation of adaptation to illusory lines. *J. Vis.*, *4*(6), 434–444.

Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychol. Sci.*, *13*(5), 416–424.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.*, *24*(5), 939–973.

Owen, A. M., Cleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science*, *313*, 1402.

Pashler, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *335*, 23–30.

Posner, M. I. (1994). Attention: The mechanisms of consciousness. *Proc. Natl. Acad. Sci. USA*, *91*(16), 7398–7403.

Purves, D., Paydarfar, J. A., & Andrews, T. J. (1996). The wagon wheel illusion in movies and reality. *Proc. Natl. Acad. Sci. USA*, *93*, 3693–3697.

Reddy, L., Reddy, L., & Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Res.*, *46*(15), 2336–2343.

Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *J. Vis.*, *4*(2), 106–117.

Rees, G., & Frith, C. (2007). *Methodologies for identifying the neural correlates of consciousness*. Oxford, UK: Blackwell.

Sacks, O. (2004). In the river of consciousness. *New York Rev. Books*, *51*, 41–44.

Schiff, N. D. (2004). *The neurology of impaired consciousness: Challenges for cognitive neuroscience*. Cambridge, MA: MIT Press.

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol. Sci.*, *15*, 720–728.

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. USA*, *94*, 3408–3413.

Siegel, J. (2000). Narcolepsy. *Sci. Am.*, *282*, 76–81.

Slater, R., Cantarella, A., Gallella, S., Worley, A., Boyd, S., Meek, J., & Fitzgerald, M. (2006). Cortical pain responses in human infants. *J. Neurosci.*, *26*, 3662–3666.

Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nat. Neurosci.*, *11*, 543–545.

Sperling, G., & Dosher, B. (1986). Strategy and optimization in human information processing. In K. R. Boff, L. Kaufman, & J. P. Thomas. (Eds.), *Handbook of perception and human performance* (pp. 1–65). New York, Wiley.

Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to reversal of action and sensation. *Neuron*, *51*, 651–659.

Sukhotinsky, I., Zalkind, V., Lu, J., Hopkins, D. A., Saper, C. B., & Devor, M. (2007). Neural pathways associated with loss of consciousness caused by intracerebral microinjection of GABA-active anesthetics. *Eur. J. Neurosci.*, *25*, 1417–1436.

Thorpe, S., Fize, D., & Marlot. C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*, 753–759.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.*, *5*, 42–72.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, *282*, 1846–1851.

Tse, P. U., Martinez-Conde, S., Schlegel, A. A., & Macknik, S. L. (2005). Visibility, visual awareness, and visual masking of simple unattended targets are confined to areas in the occipital cortex beyond human V1/V2. *Proc. Natl. Acad. Sci. USA*, *102*(47), 17178–17183.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nat. Neurosci.*, *8*, 1096–1101.

Tsuchiya, N., Koch, C., Gilroy, L. A., & Blake, R. (2006). Depth of interocular suppression associated with continuous flash suppression, flash suppression, and binocular rivalry. *J. Vis.*, *6*(10), 1068–1078.

Tulving, E. (1993). *Varieties of consciousness and levels of awareness in memory*. Oxford, UK: Oxford University Press.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends Cogn. Sci.*, *7*, 207–213.

VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *J. Cogn. Neurosci.*, *15*(2), 209–217.

Wegner, D. M. (2002). *The illusion of conscious will*. MA: Cambridge, MIT Press.

Wilke, M., Logothetis, N. K., & Leopold, D. A. (2006). Local field potential reflects perceptual suppression in monkey visual cortex. *Proc. Natl. Acad. Sci. USA*, *103*, 17507–17512.

Wolfe, J. M. (1984). Reversing ocular dominance and suppression in a single flash. *Vision Res.*, *24*, 471–478.

Zeki, S. (1998). Parallel processing, asynchronous perception, and a distributed system of consciousness in vision. *Neuroscientist*, *4*, 365–372.

Zhang, F., Wang, L. P., Adamantidis, A., de Lecea, L., & Deisseroth, K. (2007). Multimodal fast optical interrogation of neural circuitry. *Nature*, *446*(7136), 633–639.

80 Visual Awareness

GERAINT REES

ABSTRACT Human vision gives rise to subjective experience of the external world. Vision depends on signals and processing within the central nervous system, yet it is apparent that not all activity associated with vision reaches awareness. For example, the state of a single photoreceptor in the retina cannot be directly reported, even though that receptor processes sensory information that contributes to perception. Processing of visual signals can therefore occur in the absence of awareness, and a central question in the neurobiology of consciousness is which neural signals and psychological processes in the brain are correlated with visual awareness and which remain unconscious. Qualitatively and quantitatively dissociating brain activity associated with conscious and unconscious vision will reveal the neural correlates of visual awareness. This chapter will review the current state of progress in understanding the neural signals underlying human visual awareness, with a particular focus on work carried out in the last five years.

Measuring visual awareness

Visual awareness is the subjective experience of the external world mediated by the visual system. In this chapter, I will treat *awareness* and *consciousness* as interchangeable descriptions of the same subjective phenomena in humans. Visual awareness is individuated by its subjective content, which typically represents the presence, qualities (e.g., color, motion), and identity of objects in the environment. Here we will be concerned with the neural correlates of such phenomenal content, setting aside a consideration of those neural factors that contribute to the waking state and thus enable normal visual awareness (see Schiff, chapter 78, this volume, for further discussion of the neural basis of the waking state). In establishing a relationship between the contents of awareness and neural activity or psychological process, we rely on the ability of humans to report their experience, either verbally or through other means. Understanding the nature of such subjective reports is therefore important to appreciating inferential limitations in determining neural correlates of visual awareness.

RELIANCE ON SUBJECTIVE REPORTS The most common measure of visual awareness relies on observers directly reporting their experience. Such subjective reports directly capture the introspective and phenomenal, subjective, nature of consciousness. They thus have face validity. Critically, however, they depend not only on the sensitivity of an observer's brain to a visual stimulus but also on his or her decision criterion (an internal state) for reporting it. This fact renders such purely subjective measures potentially unreliable as a measure of whether a visual stimulus has reached awareness or not. For example, if observers are uncertain about their visual experience, they may adopt a conservative decision criterion and report uncertain or unclear perception of a stimulus as failure to perceive. In such circumstances erroneous inferences can potentially be made about brain activity. For example, on a trial where an observer erroneously reports unclear perception as absent perception, brain activity on that trial may be incorrectly characterized as unconscious or unaware.

OBJECTIVE APPROACHES AND THEIR VARIANTS Methods that factor out the effects of confidence and decision criterion on measures of awareness exist and are increasingly used in the study of the neural correlates of visual awareness. Compared to subjective report, objective methods such as those provided by signal detection theory furnish estimates of whether a signal is (consciously) discriminable independently of the criterion of the observer, removing the inferential problem we have described. However, such signal detection methods also have a number of disadvantages. In particular, the relatively large number of trials required to accurately determine sensitivity (d') raises the possibility that if decision criterion varies over time, even a demonstration of absent sensitivity over a number of individual trials may not be sufficient to claim that awareness was absent on every single trial.

More recently, alternative objective methods for judging the contents of visual consciousness have been put forward, such as postdecision wagering. In this method, observers make a visual discrimination and then subsequently make a cash wager about the outcome of that discrimination (Persaud, McLeod, & Cowey, 2007). In a range of tasks, observers can fail to maximize cash earnings despite making correct discriminations. This failure is taken to indicate that observers were objectively unaware of the outcome of their discrimination; otherwise, they could have wagered to maximize their earnings. However, good performance on both discrimination and wagering is taken as direct evidence of

GERAINT REES Institute of Cognitive Neuroscience and Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom

awareness of the stimuli. Such novel objective measures are intriguing, but have been criticized on conceptual grounds for failing to provide a "direct" measure of visual awareness, as wagering is taken to represent a "second-order" or "metacognitive" decision, depending on second-order information *about* the content of visual awareness, rather than that content per se (Seth, 2008). Other approaches continue to be explored, including hybrid methodologies that combine subjective report with assessment of confidence and that appear to be relatively free of response bias (Evans & Azzopardi, 2007).

The Character of Visual Awareness Signal detection theory and some theoretical approaches to consciousness suggest that mental representations, including those associated with visual awareness, are graded rather than discrete. On such an account, conscious visual percepts exist on a continuum of clarity. However, other theoretical proposals (discussed later; also see Tononi & Balduzzi, chapter 84, this volume) predict qualitative differences between unconscious and conscious processing, suggesting a discrete representation. The degree to which visual representations are graded or discrete has therefore been under increasing scrutiny. When observers are asked to rate the visibility of targets rendered difficult to see by presentation during the "attentional blink," surprisingly they use a continuous scale in an all-or-none fashion (Sergent & Dehaene, 2004). Either targets are rated as easy to identify and as clearly visible as targets presented outside the attentional blink, or they are not detected at all. This result suggests a discrete "all-or-none" phenomenological classification of stimuli as either fully visible or invisible. However, it has been argued that observers may be confused by the large number of categories and lack of clarity of definitions for categories intermediate between "visible" and "invisible" (Overgaard, Rote, Mouridsen, & Ramsoy, 2006). Indeed, when a smaller four-point scale is employed in a masked visual discrimination task, observers use the full range of the four-point scale to categorize visibility, and the probability of correct report varies systematically with the score on the visibility scale. Further work will therefore be needed to reconcile these accounts.

Whatever the method used for determining awareness, the necessity of an effective behavioral measure of whether an observer is aware or unaware of a visual stimulus is a prerequisite for valid inference about the neural correlates of consciousness. Without such a measure, experiments that measure neural activity associated with visual stimulation cannot make secure inferences about neural correlates of conscious or unconscious processing. This is a particular problem for experiments using animal models of human consciousness; though ingenious behavioral measures have been described (Cowey & Stoerig, 1997), they often require lengthy training. Indeed, it has been proposed that identifying hallmarks of visual awareness in nonhuman animals might need to rely on strategies other than behavioral report, such as searching for evolutionary homologies in anatomical substrate and measurement of neural correlates of conscious states (Edelman, Baars, & Seth, 2005).

Relating visual awareness to brain activity

The ability to characterize visual awareness behaviorally opens up the possibility of relating such reports to brain activity. Determining the neural correlates of the contents of visual awareness requires dissociation of such activity from activity associated with merely unconscious aspects of perception or action (see also Lau, chapter 83, this volume). The most empirically tractable way of controlling for these unwanted effects is to keep them constant, and it has been proposed (Frith, Perry, & Lumer, 1999) that it is useful to distinguish three kinds of neural activity:

1. Neural activity associated with the contents of visual awareness
2. Neural activity associated with changes in a visual stimulus, in the absence of changes in visual awareness
3. Neural activity associated with visually guided behavior, in the absence of changes in visual awareness

Such an approach can be used to elucidate a taxonomy classifying the types of experimental paradigms relevant to these three different types of neural activity, while remembering that whether they can be distinguished is ultimately an empirical question. Table 80.1 illustrates some examples of neuroimaging and behavioral experiments in the domain of visual awareness that fit such a taxonomy.

Having established how reports of visual experience can be assessed, and appropriate experimental designs used to dissociate the neural correlates of conscious and unconscious vision, we can proceed to consider the neural substrates of each in turn.

Characterizing the unconscious homunculus

Several decades of behavioral research have led to significant understanding of the potential theoretical and empirical difficulties associated with studying unconscious processing (Kouider & Dehaene, 2007). Although the debate over the existence of unconscious perception continues (Hannula, Simons, & Cohen, 2005), most authors accept the occurrence of implicit visual perception, where unconscious visual processing can alter behavior. A significant body of brain-imaging studies now shows that many regions of human visual cortex can be activated by visual stimuli that do not reach awareness (see figure 80.1).

Activity can be identified in the absence of visual awareness at the earliest retinotopic cortical stages of visual pro-

TABLE 80.1

Examples of experimental paradigms for studying the neural correlates of visual awareness

In each category two examples are given of experimental paradigms that have been used to identify neural correlates of visual awareness, visual stimulation, or visually guided behavior.

	Examples of Paradigms	
Subjective experience changes, visual stimulation and/or behavior remain constant	Neural correlates of binocular rivalry (e.g., Haynes, Deichmann, & Rees, 2005).	Neural correlates of change awareness (Beck, Rees, Frith, & Lavie, 2001)
Visual stimulation changes, subjective experience remains constant	Adaptation under conditions of visual crowding (e.g., He, Cavanagh, & Intriligator, 1996)	Neural correlates of visual masking (e.g., see Macknik & Martinez-Conde, chapter 81, this volume)
Behavior changes, subjective experience remains constant	Neural correlates of correct discrimination without awareness (e.g., Sahraie et al., 1997)	Unconscious priming of cognitive control (e.g., Lau & Passingham, 2007)

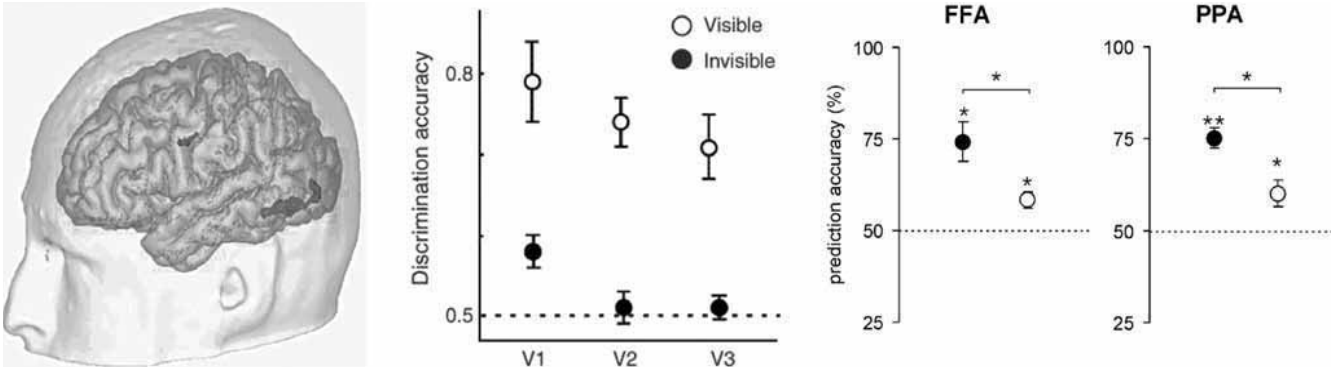


FIGURE 80.1 Activation of sensory cortices by invisible stimuli. *Left panel*: Masked and invisible words nevertheless evoke activation (shown in orange, superimposed on an anatomical image of the brain) of the fusiform gyrus. See Dehaene et al. (2001) for further details. *Middle panel*: Activity measured using BOLD contrast functional MRI in human V1–V3 can be used to discriminate the orientation (right or left tilted) of a grating stimulus. Open symbols representing mean decoding accuracy for a group of subjects (error bars, one SE) for visible stimuli; closed symbols for similarly oriented stimuli rendered invisible by masking. The orientation of these invisible stimuli can still be discriminated at a rate significantly better than chance in human V1. See Haynes and Rees (2005a) for further details. *Right panel:* Performance of support-vector-machine (SVM) classifiers for pairwise classification of face and house presentations from fusiform face area (FFA) and parahippocampal place area (PPA). Average prediction accuracies across participants for visible faces versus houses are denoted by filled circles (±SEM) and for invisible faces versus houses by empty circles. The dotted lines denote chance level (50%). $*p < 0.05$; $**p < 0.005$; n.s. = non significant ($p > 0.1$). Above-chance performance for invisible stimuli indicates that information is still present in higher visual areas sufficient to discriminate stimulus category. See Sterzer, Haynes, and Rees (2008) for further details. (See color plate 97.)

cessing. For example, when a simple achromatic disc is flashed briefly, activity is elicited in the corresponding retinotopic location of primary visual cortex (V1) even when the target is rendered completely invisible by a surrounding mask (Haynes, Driver, & Rees, 2005). This activation of visual cortex by stimuli that cannot be accurately reported continues to higher stages of processing. For example, activation of functionally specialized areas is consistently observed for masked words, faces, and objects that observers do not report seeing. Masked and unreported words can activate the "temporal word form area" (Dehaene et al., 2001), and activation associated with words that are not perceived can extend to left-hemisphere language regions (Diaz & McCarthy, 2007). Dichoptically masked object and face stimuli can nevertheless activate functionally specialized areas of ventral visual cortex (Moutoussis & Zeki, 2002), and category selectivity can also be identified in responses of these areas to stimuli rendered invisible through interocular suppression (Sterzer, Haynes, & Rees, 2008). Such observations are not restricted to different types of visual masking, as unconscious activation of the ventral visual pathway during the "attentional blink" can reflect object category (Marois, Yi, & Chun, 2004) and semantic analysis of visually presented words (Luck, Vogel, & Shapiro, 1996). Similarly, visual motion rendered invisible through "crowding" can still activate V5/MT (Moutoussis & Zeki, 2006). Unconscious activation also extends to cortical areas considered part of the dorsal stream of visual processing; images of tools rendered invisible by continuous flash suppression nevertheless activate the human dorsal stream (Fang & He, 2005).

Such neuroimaging evidence converges with electrophysiological studies showing that electrical potentials associated with higher levels of visual processing can be elicited by invisible stimuli. For example, the N400 potential—thought to reflect semantic processing—can be modulated by unconsciously perceived masked prime words (e.g., Luck et al., 1996; though see Kouider & Dehaene, 2007, for a thorough review of evidence for and against unconscious semantic priming), and such effects persist even when prime visibility is carefully controlled to ensure that masked primes are indeed invisible (Kiefer & Brendel, 2006). Convergence between fMRI and electrophysiological measures is not confined to words; both blood oxygenation level dependent (BOLD) signals and event-related potentials in the ventral visual pathway are modulated by subliminally presented face stimuli, indicating unconscious face processing (Henson, Mouchlianitis, Matthews, & Kouider, 2008; Kouider, Dehaene, Jobert, & Le Bihan, 2007).

Many of these neuroimaging studies rely on subjective behavioral measures of awareness (see previous discussion) and have been criticized for not eliminating the possibility of low-confidence conscious perception confounding trials where masked stimuli were not reported (Hannula et al., 2005). Nevertheless they show striking convergence with the rather small number of experiments that assess awareness that employ "objective" behavioral measures or rating scales (discussed earlier). For example, even when participants' visibility ratings do not deviate from the lowest value on a scale, occipitotemporal event-related potentials can be evoked by an invisible word (Sergent, Baillet, & Dehaene, 2005). Taken together, they suggest that under conditions that render stimuli subjectively or objectively invisible, a substantial degree of processing continues in visual cortex, including higher stages of visual processing.

Subcortical pathways can also show activation in response to subjectively invisible and unreported emotional visual stimuli. For example, the amgydala responds selectively to fearful faces under conditions of masking and binocular suppression, even when observers are unable to report their presence (e.g., Morris, Ohman, & Dolan, 1999; Williams, Morris, McGlone, Abbott, & Mattingley, 2004). Notably, however, these results were obtained using "subjective" measures of behavioral report, and as individuals show large interindividual differences in sensitivity to emotional faces (Pessoa, Japee, & Ungerleider, 2005) there are likely to be large differences between objective and subjective thresholds for awareness of emotional stimuli, a fact which similarly places constraints on the strength of the conclusions that can be drawn about unconscious processing in subcortical pathways (see Pessoa, 2005, for a thorough discussion of these issues).

Many of the studies described here use masking or interocular suppression approaches to render stimuli invisible. Neuronal populations signals elicited by such masked or suppressed stimuli are generally substantially lower than for the equivalent unmasked stimuli that reach awareness. Such observations have led to theoretical claims (Zeki, 2008) that conscious perception of a given visual attribute resides in the extrastriate area specialized for that attribute (e.g., area MT/V5 for motion or area V4 for color). While this topic will be discussed further in this chapter and theoretical approaches discussed elsewhere (see Tononi & Balduzzi, chapter 84, this volume), it is sufficient to note at this point that several studies using paradigms other than masking show that invisible stimuli can nevertheless elicit population neuronal activity or event-related potentials equivalent in amplitude to visible stimuli. The degree to which signals associated with invisible stimuli are attenuated therefore appears to depend on the experimental paradigm rather than visibility per se, though this question merits further investigation.

Probing unconscious vision with multivariate pattern analyses

Neurons in early visual cortex are sensitive to a number of visual features, such as orientation and direction of motion. It is well established that orientation-selective aftereffects can result from exposure to grating stimuli that are too fine to be consciously perceived (He & MacLeod, 2001), suggesting that orientation-selective but unconscious activation of visual cortex is possible. Direct physiological measurement of such unconscious feature-selective processing in human V1 has proven elusive because of the relatively low spatial resolution (several millimeters) of functional neuroimaging methods compared to the size of orientation columns in visual cortex (hundreds of microns). However, it has recently become possible to use fMRI even at conventional resolutions (typically, voxels measuring $3 \times 3 \times 3$ mm) to obtain a direct measure of orientation-selective processing in V1 (Haynes & Rees, 2005a; Kamitani & Tong, 2005). Many individual voxels in V1 show subtle but reproducible biases in their activity when differently oriented stimuli are presented to the experimental subject. This result may reflect biased sampling by the (relatively) low-spatial-resolution voxels of the underlying columnar organization of orientation-specific neuronal populations, resulting from an uneven distribution of different orientation specificities across the cortical surface (for more detailed explanations, see Kamitani & Tong, 2005, or Haynes & Rees, 2005a). Importantly, this information can be efficiently accumulated across the whole of V1 using multivariate pattern recognition analyses (for reviews see Haynes & Rees, 2006; Norman, Polyn, Detre, & Haxby, 2006). Such multivoxel pattern analysis can successfully predict which one of two oriented stimuli a participant is viewing, even when masking renders that stimulus invisible (Haynes & Rees, 2005a). This finding indicates the presence

of feature-selective processing in early visual cortex, even for invisible stimuli. At higher levels of the visual system, multivariate pattern analyses reveal unconscious representation of category-selective information about face and house stimuli rendered invisible by continuous flash suppression (Sterzer et al., in press). Such analyses thus open up the possibility of probing different levels of neuronal representation associated with conscious or unconscious processing.

Awareness making a difference

Visual stimuli can therefore elicit considerable processing outside awareness, and this can influence subsequent behavior. Why might this statement be true? Crick and Koch (2000) have argued that the brain may consist of a series of specialized unconscious "zombie" systems that can control behavior on the basis of visual signals but in the absence of awareness. Such "zombie systems" raise the question of why the brain does not simply consist of a series of such specialized systems. Crick and Koch (2000) argue that such an arrangement would be inefficient in circumstances where many such systems are required, such as for complex organisms capable of generating multiple behaviors such as humans. In such situations the processing of unconscious systems might be better used to produce a single more complex representation that is available to make a choice among different but possible plans for action.

To understand such a claim we need to establish whether differences exist in the pattern and character of neural activity associated with stimuli that reach awareness compared to those that do not. Studies that have examined this question typically contrast situations where physically identical visual stimulation either gives rise to awareness or does not (Frith et al., 1999). Visual awareness is associated with changes not only in the level and timing of neural activity, but also in the cortical areas that show changes in activity.

Level of Neural Activity A common but not invariable finding is that awareness of a particular visual feature or object in the environment is associated with enhancement of neuronal signals in functionally specialized regions of visual cortex whose neuronal specificities represent that stimulus attribute or object category. At the anatomically earliest stages of visual processing, signals in V1 scale linearly with the magnitude of change in retinal illumination, as do subjects' subjective ratings of the perceived brightness of the stimuli (Haynes, Lotto, & Rees, 2004). Such a close correspondence between fMRI signals and phenomenal perception is consistent with a role for these areas in representing conscious contents. More direct evidence has been provided by studies of sensory stimulation at perceptual threshold. For simple grating stimuli, a stimulus that is successfully detected by a subject evokes significantly greater activity in V1 compared to identical stimuli that do not reach awareness (Ress & Heeger, 2003). Importantly, the change in overall level of visual cortical activity associated with conscious (versus unconscious) visual stimulation is not identified in all paradigms (discussed earlier). Visual awareness does not therefore invariably depend on a modulation of activity in functionally specialized regions of visual cortex. A certain level of neuronal activity may be necessary but not sufficient for awareness.

Timing of Neural Activity Conscious detection of threshold-level stimuli is associated with very early differential electrical signals at posterior electrodes (Pins & Ffytche, 2003), suggesting that changes in conscious contents can be associated with activity that is both temporally and anatomically early in processing. However, it is important to note that it is not necessarily the case that conscious and unconscious processing can always be distinguished early in time. In some experimental situations, conscious and unconscious processing can only be dissociated at a much later stage, after several hundred milliseconds (Sergent et al., 2005; Vogel, Luck, & Shapiro, 1998). Neither need it be the case that such late divergence does not reflect V1 involvement, as such temporally delayed correlates of conscious perception can be identified in monkey V1 (Super, Spekreijse, & Lamme, 2001), supporting theoretical suggestions that conscious perception correlates not with the "feedforward" sweep of information processing following stimulation processing but rather with later feedback or recurrent signals, perhaps to V1 (Lamme & Roelfsema, 2000).

Involvement of Early Visual Cortex In V1, retinotopic activity can reflect conscious perception of illusory features. When a moving grating is divided by a large gap, observers report seeing a moving "phantom" in the gap, and there is enhanced activity in the locations in early retinotopic visual cortex corresponding to the illusory percept (Meng, Remus, & Tong, 2005). Moreover, when phantom-inducing gratings are paired with competing stimuli that induce binocular rivalry, spontaneous fluctuations in conscious perception of the phantom occur together with changes in early visual activity. Similarly, V1 activation can be found on the path of apparent motion (Muckli, Kohler, Kriegeskorte, & Singer, 2005) and is associated with strengthened feedback connections to that retinotopic location from cortical area V5/MT (Sterzer, Haynes, & Rees, 2006). Finally, when two objects subtending identical angles in the visual field are made to appear of different sizes using three-dimensional context, the spatial extent of activation in the V1 retinotopic map reflects the perceived rather than actual angular size of the objects (Murray, Boyaci, & Kersten, 2006). These data thus show that either the level or spatial extent of V1 activation can correspond to the perceived phenomenal properties of the visual world independently of visual stimulation. This

finding extends to cross-modal influences on visual perception. Irrelevant auditory stimulation can lead to illusory perception of a single flash as two flashes, and primary visual cortex shows enhanced activity compared to physically identical stimulation that is perceived correctly (Watkins, Shams, Tanaka, Haynes, & Rees, 2006). Moreover, these alterations in the contents of visual consciousness are associated with very early modulation of MEG responses over posterior occipital sensors (Shams, Iwaki, Chawla, & Bhattacharya, 2005). Responses in human V1 can therefore be altered by sound, and can reflect subjective perception rather than the physically present visual stimulus. Although neuronal population activity in V1 can show a close relationship with the contents of visual awareness, the earlier discussion of data showing that neuronal populations in V1 can also be activated in the absence of awareness indicates that V1 activity alone cannot be sufficient for awareness. Either the precise character of V1 activity or involvement of other brain areas must be required.

Involvement of Higher Visual Cortices V1 is not the only area in the human visual pathway whose activity can reflect the contents of consciousness. Visually presented objects can be made difficult to identify by degrading them, and in such circumstances occipitotemporal activity shows a close correlation with recognition performance (Grill-Spector, Kushnir, Hendler, & Malach, 2000). Similarly, conscious detection of changes in a visually presented object is associated with enhanced activity in ventral visual cortex (Beck et al., 2001). There are many other examples of similarly close correspondence between the level of activity in the ventral visual pathway and conscious perception. For example, patients with schizophrenia who experience visual hallucinations show activity in modality-specific cortex during hallucinatory episodes (Silbersweig et al., 1995). Similarly, patients with damage to the visual system who experience hallucinations with specific phenomenal content show activity in functionally specialized areas of visual cortex corresponding to the contents of their hallucinations (Ffytche et al., 1998). In healthy volunteers, visual imagery activates category-specific areas of visual cortex (O'Craven & Kanwisher, 2000) and category-selective neurons in the human medial temporal lobe (Kreiman, Koch, & Fried, 2000). Contingent aftereffects based on color or motion lead to activation of either V4 (Barnes et al., 1999; Sakai et al., 1995) or V5/MT (He, Cohen, & Hu, 1998; Tootell et al., 1995), respectively, and the time course of such activation reflects phenomenal experience. Perception of illusory or implied motion in a static visual stimulus is associated with activation of V5/MT (e.g., Zeki, Watson, & Frackowiak, 1993), whereas perception of illusory contours activates extrastriate cortex (Hirsch et al., 1995). Differential activity in word-processing areas is present when subjects are consciously aware of the meaning of visually presented words and absent when they are not (Rees, Russell, Frith, & Driver, 1999).

Some of the most popular paradigms for studying the neural correlates of visual awareness are bistable phenomena such as binocular rivalry. When dissimilar images are presented to the two eyes, they compete for perceptual dominance so that each image is visible in turn for a few seconds while the other is suppressed (see figure 80.2). This binocular rivalry is associated with relative suppression of local, eye-based representations that can also be modulated by high-level influences such as perceptual grouping (see Tong, Meng, & Blake, 2006, for a review). Because perceptual transitions between each monocular view occur spontaneously without any change in the physical stimulus, neural correlates of consciousness may be distinguished from neural correlates attributable to stimulus characteristics. All stages of visual processing show such activity changes associated with rivalrous fluctuations. For example, even at the earliest subcortical stages of visual processing, signals recorded from the human lateral geniculate nucleus (LGN) exhibit fluctuations in activity during binocular rivalry (Haynes, Deichmann, et al., 2005; Wunderlich, Schneider, & Kastner, 2005). Primary visual cortex shows a similar pattern of changes in activity correlated with changes in the contents of consciousness (S. Lee & Blake, 2002; S. Lee, Blake, & Heeger, 2005; Polonsky, Blake, Braun, & Heeger, 2000; Tong & Engel, 2001). In general (though see Tong & Engel, 2001) such fluctuations in activity are about half as large as those evoked by nonrivalrous stimulus alternation. This difference indicates that the suppressed image during rivalry undergoes a considerable degree of unconscious processing. Finally, further along the ventral visual pathway, responses in fusiform face area (FFA) during rivalry are larger than those in V1, and equal in magnitude to responses evoked by nonrivalrous stimuli (Tong, Nakayama, Vaughan, & Kanwisher, 1998). This finding suggests that neural competition during rivalry has been resolved by these later stages of visual processing, and activity in FFA thus reflects the contents of consciousness rather than the retinal stimulus. However, such an account is inconsistent with the finding that binocularly suppressed faces can nevertheless still activate the FFA (Moutoussis & Zeki, 2002) and with the recent demonstration of category-selective signals in these areas for binocularly suppressed face or house stimuli (Sterzer et al., in press).

Spatial Patterns of Activity in Ventral Visual Cortex As with unconscious processing, pattern-based decoding approaches have recently been applied to fMRI signals from human ventral visual cortex. These techniques have the

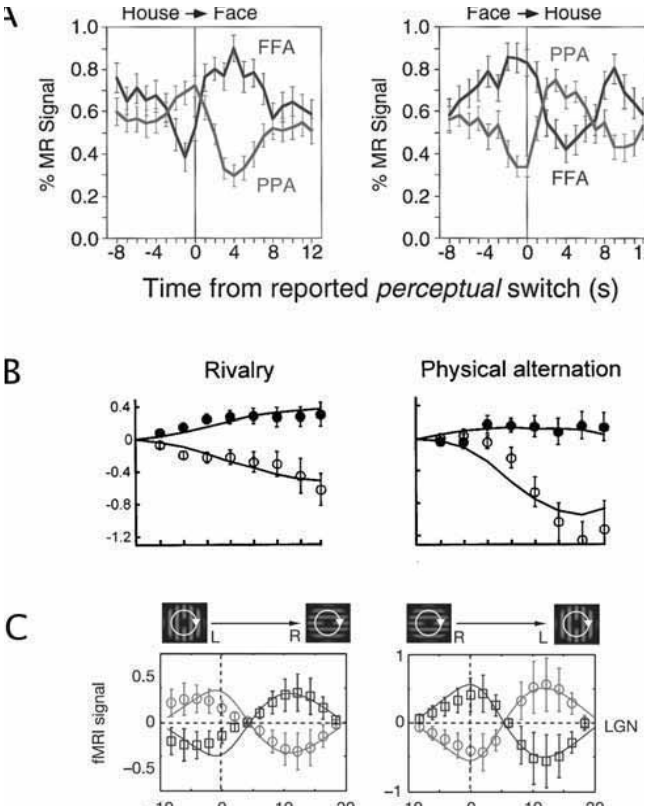


Figure 80.2 Fluctuations in activity in visual pathways associated with visual awareness during binocular rivalry. (*A*) Fusiform face area. Activity measured using functional MRI from human fusiform face area (FFA) and parahippocampal place area (PPA) is plotted as a function of time relative to a perceptual switch from house to face (left panel) or face to house (right panel). It is apparent that activity in the FFA is higher when a face is perceived during binocular rivalry than when it is suppressed, and activity in the PPA is similarly higher when a house is perceived than when it is suppressed. For further details see Tong, Nakayama, Vaughan, and Kanwisher (1998). (*B*) Binocular rivalry in primary visual cortex (V1). Activity measured using fMRI from human primary visual cortex is plotted as a function of time after a perceptual switch where the subsequent perception is of a high-contrast stimulus (solid symbols) or low-contrast stimulus (open symbols). The left-hand panel plots activity following a perceptual switch due to binocular rivalry, while the right-hand panel plots activity following a deliberate physical switch of monocular (nonrivalrous) stimuli. V1 activity therefore corresponds to perception during binocular rivalry, and the amplitude changes are similar to those seen during physical alternation of corresponding monocular stimuli. For further details see Polonsky, Blake, Braun, and Heeger, (2000). (*C*) Rivalry in the lateral geniculate nucleus (LGN). Activity measured using fMRI is plotted as a function of time for voxels in the LGN selective for left-eye stimuli (red symbols) or right-eye stimuli (blue symbols) around the time (vertical dotted line) of a perceptual switch between left- and right-eye views (left panel) or right- and left-eye views (right panel). Reciprocal changes in signal in the different eye-selective voxels as a function of perceptual state can be readily seen. For further details see Haynes, Deichmann, and Rees (2005). (See color plate 98.)

potential to reveal qualitative differences in neural processes underlying conscious and unconscious processing in a cortical area. Orientation and direction of motion of simple visual stimuli can be decoded (Haynes & Rees, 2005a; Kamitani & Tong, 2006), plus the identity of more complex objects (Haxby et al., 2001). If subjects are asked to attend one of two overlapping orientations or motion directions, then patterns of activity in early visual cortex can be used to predict which one is attended (Kamitani & Tong, 2005, 2006). Moreover, the local spatial pattern of activity in early retinotopic visual cortex can be used to dynamically decode and accurately predict perceptual state during binocular rivalry over extended periods of time (Haynes & Rees, 2005b). These data suggest that reliable decoding of the subjective contents of consciousness, at least under controlled viewing conditions, may be a realistic prospect. Moreover, the ability to provide information about the underlying specificities of the neuronal populations may now provide a future basis for characterizing whether conscious and unconscious stimuli elicit different types of activity in a single cortical region.

Involvement of Frontal and Parietal Cortex Although visual cortex plays a central role in representing the contents of visual consciousness, there is now considerable evidence that activity in frontal and parietal cortex is strongly correlated with changes in the contents of visual awareness (see figure 80.3). For example, activity during transitions in binocular rivalry, as well as other forms of bistable perception, is time-locked to frontal and parietal cortex activity (Kleinschmidt, Buchel, Zeki, & Frackowiak, 1998; Lumer, Friston, & Rees, 1998; Sterzer, Russ, Preibisch, & Kleinschmidt, 2002). Strikingly, frontal and parietal activity is also associated with spontaneous changes in the contents of consciousness in a variety of perceptual paradigms, such as stereo pop-out (Portas, Strange, Friston, Dolan, & Frith, 2000), the perception of fragmented figures (Eriksson, Larsson, Riklund Ahlstrom, & Nyberg, 2004), the detection of change in a visually presented object (Beck et al., 2001), conscious perception of flicker (Carmel, Lavie, & Rees, 2006), and successful conscious identification of visually masked words (Dehaene et al., 2001). Electrical activity over parietal sensors is associated with the detection of a simple threshold-level stimulus (Pins & Ffytche, 2003). Moreover, changes in the contents of consciousness during bistable perception are associated with distributed changes in synchronous electrical oscillations measured on the scalp (Srinivasan, Russell, Edelman, & Tononi, 1999; Struber & Herrmann, 2002; Tononi, Srinivasan, Russell, & Edelman, 1998).

Importantly, frontoparietal activity is not simply associated with the requirement that observers report their experience. When binocular rivalry occurs in the absence of

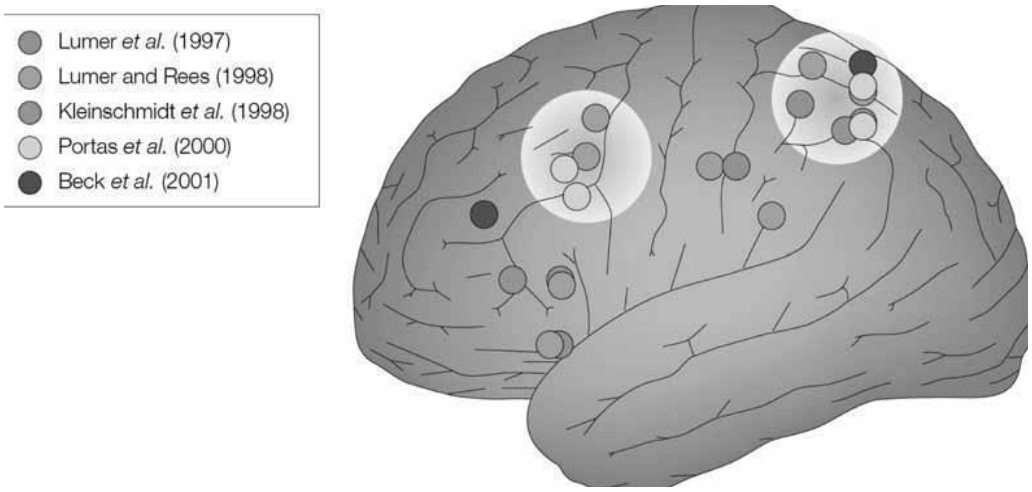


FIGURE 80.3 Parietal and prefrontal correlates of visual awareness. Foci of parietal and prefrontal activity measured using functional MRI and associated with switches in the contents of consciousness independent of changes in physical stimulation are plotted on an anatomical brain image in a standard stereotactic space. Studies shown identify the neural correlates of perceptual switches during rivalry (Lumer, Friston, & Rees, 1998; Lumer & Rees, 1999), during bistable perception generally (Kleinschmidt, Buchel, Zeki, & Frackowiak, 1998), associated with stereo pop-out (Portas, Strange, Friston, Dolan, & Frith, 2000), or associated with change detection (Beck, Rees, Frith, & Lavie, 2001). Clustering of activated foci (white circles) are apparent in superior parietal and dorsolateral prefrontal cortex. (See color plate 99.)

behavioral reports, there is a close coupling between activity in early visual cortical areas representing the rivaling stimuli and multiple regions of frontal and parietal cortex previously associated with the report of conscious transitions (Lumer & Rees, 1999). Such frontal and parietal involvement in rivalry therefore appears independent of the requirement to make motor reports (though also see Lau, chapter 83, this volume).

INVOLVEMENT OF SPECIFIC PSYCHOLOGICAL PROCESSES Demonstrating that activity in specific brain areas is associated with visual awareness does not determine which psychological processes might give rise to awareness. Nevertheless, the common association of frontal and parietal activation with awareness has led to speculation that selective attention may be both necessary and sufficient for awareness of a stimulus to arise (e.g., Block, 2007). Indeed, failure to allocate attention to stimulus processing can lead to stimuli neither being perceived nor eliciting activity associated with their identity (Rees et al., 1999). However, more recently both behavioral and neuroimaging evidence converge to indicate that selective attention can facilitate processing and activity evoked by an invisible target (Bahrami, Lavie, & Rees, 2007; Kentridge, Heywood, & Weiskrantz, 2004; Sumner, Tsai, Yu, & Nachev, 2006; for a review see Koch & Tsuchiya, 2007). Even if attention is necessary, it therefore cannot be a sufficient precondition for visual awareness. Indeed, magnetoencephalographic signals show distinct and independent neural correlates of visual awareness and spatial attention at different frequencies in the gamma (30–150 Hz) range (Wyart & Tallon-Baudry, 2008). In contrast to attention, the potential role of other psychological processes, such as memory, in modulating cortical activity associated with visual awareness has come under rather less scrutiny but may represent a potentially fertile area for future study. For example, mnemonic signals such as those observed during interruptions in bistable perception might play a fundamental role in stabilizing conscious visual percepts (Sterzer & Rees, 2008).

Causal factors and pathologies of vision

Noninvasive approaches such as fMRI and EEG/MEG in healthy human volunteers can reveal the correlation between neural activity and the contents of visual awareness. However, the existence of such an association does not imply a causative relationship between the two. In order to do so, neural activity must be manipulated either experimentally or through studying individuals with brain damage.

TRANSIENT DISRUPTION OF CORTICAL FUNCTION Transcranial magnetic stimulation (TMS) can be used to transiently stimulate or disrupt local cortical processing. In both healthy volunteers and individuals who are blind due to retinal damage, TMS applied to visual cortex can elicit phosphenes. This ability demonstrates that retinal activity is not necessary for conscious visual experience. Similarly, visual experiences significantly more complex than mere phosphene perception can be elicited by direct electrical stimulation of the ventral visual cortex in humans (H. Lee, Hong, Seo, Tae, & Hong,

2000), suggesting that activity in the lateral geniculate nucleus may also not be necessary for visual awareness. Whether activity in primary visual cortex is necessary is more controversial. For example, TMS applied to visual cortex does not elicit phosphenes when blindness results from damage to V1 (Cowey & Walsh, 2000). However, even observers with V1 damage can experience phenomenal percepts (Weiskrantz, Rao, Hodinott-Hill, Nobre, & Cowey, 2003), indicating that V1 is necessary for some types of visual percepts. One possibility is that a specific aspect of activity in V1, such as its character or timing (discussed earlier), plays an important role. Consistent with this, awareness of motion is impaired if feedback signals from V5/MT to V1 are disrupted by TMS (Pascual-Leone & Walsh, 2001; Silvanto, Cowey, Lavie, & Walsh, 2005). Similarly, using TMS to disrupt processing of a metacontrast mask presented after a target can lead to unmasking and corresponding visibility of the original target (Ro, Breitmeyer, Burton, Singhal, & Lane, 2003). These data suggest that temporally late signals in V1 representing feedback from other ventral visual (or higher cortical) areas may be required for awareness. TMS also provides direct evidence for a role of frontal and parietal cortices in visual awareness, as detection of visual change is impaired when frontal or parietal cortex is transiently disrupted (Beck, Muggleton, Walsh, & Lavie, 2006; Turatto, Sandrini, & Miniussi, 2004).

Hemianopia and Blindsight Damage to primary visual cortex is typically associated with a lack of awareness for stimuli presented at corresponding points in the visual field. However, when patients with such cortical damage are asked to guess properties of stimuli that they deny being able to see when they are presented in a scotoma, a number show residual visual capacity in their blind field. These patients are able to perform certain discriminations and localizations better than chance in the acknowledged absence of awareness. This ability has become known as blindsight. The prevalence of such a phenomenon is unclear, but when systematic forced-choice testing with stimuli appropriate for the spatial and temporal characteristics of the residual vision found in blindsight has been applied to patients with visual field defects resulting from occipital damage, the majority demonstrate blindsight (Sahraie et al., 2006). Individuals with blindsight can still experience some types of nonveridical percept in their blind field, such as afterimages from unseen stimuli (Weiskrantz et al., 2003), but the absence of awareness associated with V1 damage points to a significant role for V1 in visual awareness. Indeed, ingenious exclusion tasks (where the participants are required to report stimuli that are *not* shown) have been used to demonstrate that blindsight is unlike normal visual awareness and does not simply reflect unreliable subjective reports (Persaud & Cowey, 2008). Many different hypotheses have been proposed for why V1 damage leads to blindness, from the inability of weaker signals from extrageniculostriate pathways to engender awareness to V1 damage interfering with particular pathways or timing of signals necessary for awareness. The recent successful characterization of anatomical pathways underlying the residual vision in one blindsighted observer (Bridge, Thomas, Jbabdi, & Cowey, 2008) points to the need to understand such deficits in the context of the connectivity of the visual system.

Visual Neglect and Visual Extinction Damage to frontal and parietal cortex is commonly associated with visual neglect and visual extinction, where patients do not perceive or respond to any type of visual stimulus placed in one half of the visual field (Driver & Mattingley, 1998). This deficit in conscious visual representation is consistent with a general involvement of these structures in representing many different types of conscious contents. Activation of visual cortex or subcortical structures by visual stimuli is insufficient for awareness in visual neglect (Driver, Vuilleumier, Eimer, & Rees, 2001). When visual stimuli are presented to patients with visual extinction, areas of both primary and extrastriate visual cortex that are activated by a seen left-visual-field stimulus are also activated by an unseen and extinguished left-visual-field stimulus (Rees et al., 2000, 2002; Vuilleumier et al., 2001) and also associated with enhanced covariation of activity between visual cortical areas representing the visual stimulus and undamaged parietal and prefrontal regions (Vuilleumier et al., 2001). The unconscious processing of an extinguished face stimulus extends even to face-selective cortex in the fusiform face area (Rees et al., 2002). These data confirm that activation of visual cortex is insufficient to result in awareness and strongly support a role for frontal and parietal structures in visual awareness (see figure 80.4).

Empirical and theoretical integration

The last decade of empirical study of the relationship between brain activity and visual awareness has given rise to an enormous amount of data, out of which certain common themes may be emerging. First, it appears that visual stimuli that do not reach awareness nevertheless undergo extensive processing in both dorsal and ventral visual pathways, although the nature and extent of processing at higher levels remain under debate. Second, neuropsychological and neuroimaging evidence strikingly converges to suggest an association of frontal and parietal activity with awareness, particularly of changes in the visual environment. Whether this reflects a particular cognitive process (such as attention) remains a matter of debate. Third, changes in coupling between different brain areas and in particular between frontal and parietal cortex with visual cortex appear to play

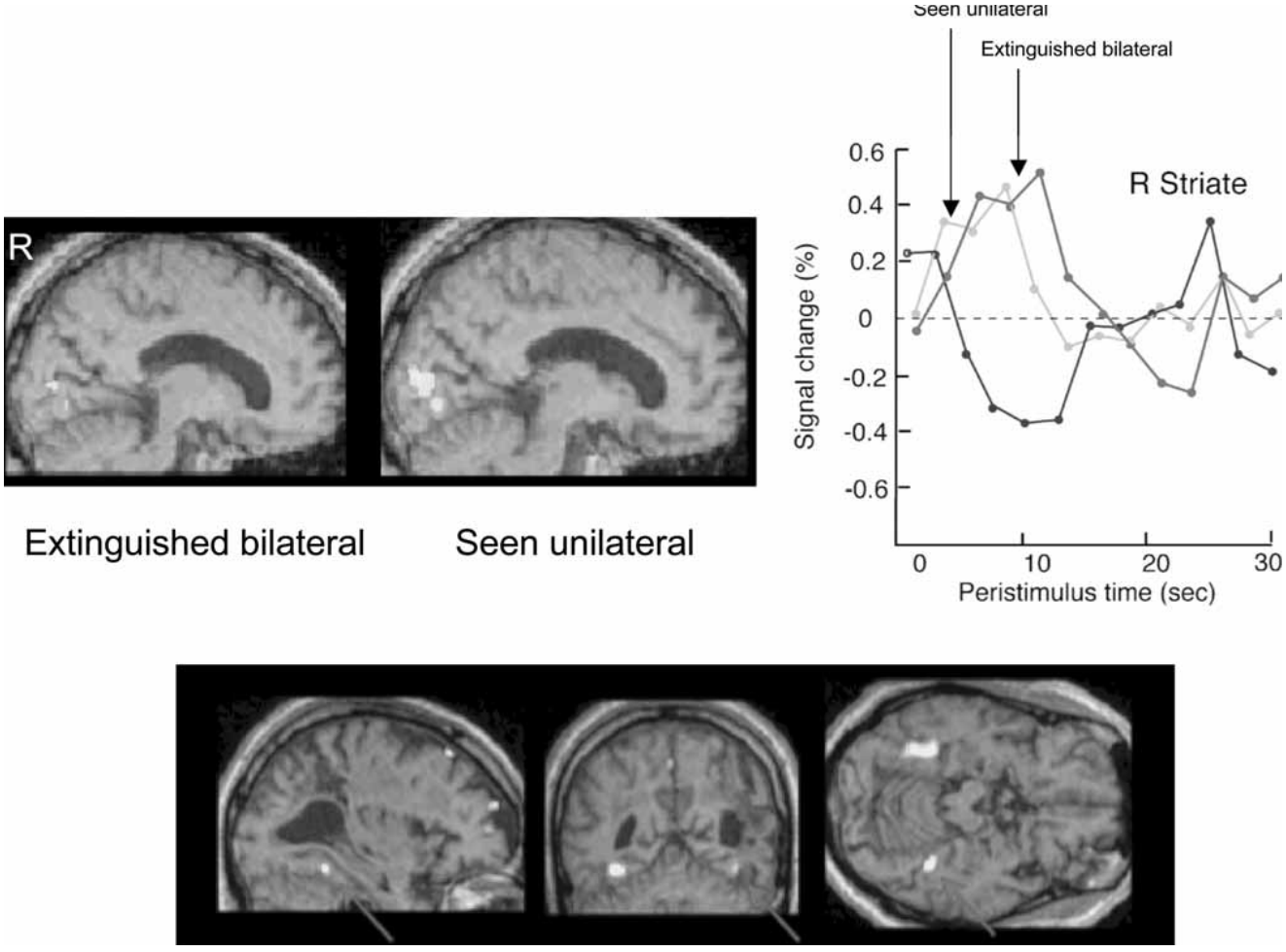


FIGURE 80.4 Activity in ventral visual cortex is not sufficient for awareness. *Upper left*: Activity evoked by an unseen and extinguished left-visual-field stimulus (see text for description of visual extinction) in striate and extrastriate cortex in a patient with parietal neglect. Differences in activity comparing bilateral extinguished with unilateral right stimulation is overlaid on two sagittal slices of a T1-weighted anatomical scan. For further details see Rees and colleagues (2000). *Upper right*: BOLD activity from the right striate (V1) focus of activation shown in the left panel is plotted as a function of peristimulus time for bilateral extinguished stimuli, unilateral left-visual-field stimuli, and unilateral right-visual-field stimuli. Note the similarity of the BOLD time courses for the bilateral extinguished stimuli (in which a stimulus is present in the left visual field but not reported by the individual with parietal extinction) and for the left unilateral stimulation (in which a stimulus is both present in the left visual field and reported). This indicates that stimuli that do and do not reach awareness may produce similar levels of activation in similar cortical locations, in this case following parietal damage. *Lower panel*: Activity in the fusiform face area evoked by a face (versus a house) stimulus presented in the neglected left hemifield of a patient with parietal neglect and left visual extinction. Thus, after parietal damage, activation in the ventral visual pathway for unseen stimuli can be sufficient to distinguish the category of stimulus presented. See Rees and colleagues (2002) for further details. (See color plate 100.)

an important role in stimuli becoming reportable and merit further investigation.

These conclusions from the study of visual awareness bear upon theoretical models of consciousness in a number of important ways (see Tononi & Balduzzi, chapter 84, and Block, chapter 77, this volume). The involvement of areas outside the classical visual pathway is not easy to explain for theoretical approaches that posit that conscious perception of a given visual attribute resides in the extrastriate area specialized for that attribute. While the contents of visual awareness may be determined by the location of neural activity in functionally specialized visual areas, this activity does not seem to be enough. Interaction with other areas, particularly parietal and prefrontal cortex, seems to be necessary. Theoretical approaches to consciousness as a whole that propose an interaction between processes specific to particular sensory domains with a "global workspace" (e.g., Dehaene, Kerszberg, & Changeux, 1998) or that highlight the role of interactions between distributed neuronal systems in general (Tononi & Edelman, 1998) appear to fit the empirical data more closely. Theoretical considerations are developed more fully elsewhere (see Tononi & Balduzzi).

Conclusion

The most parsimonious account of currently available data is that awareness of particular attributes or objects in the visual environment is contingent on the presence of an activated representation in primary visual cortex and ventral visual pathways corresponding to the attributes represented in awareness, together with activity in specific parietal (and perhaps) prefrontal structures. Activity in such a distributed network may be necessary for accurate report, which is at the heart of our attribution of consciousness to individuals. The challenge for the future is both conceptual and empirical. Conceptual and theoretical advances are required to make sense of the mass of the empirical data and integrate findings from multiple domains of inquiry. Empirical work should more precisely specify the interactions between and potential causal role for different brain areas. At the same time, a full understanding of the neural correlates of visual awareness will require detailed knowledge of the relationship between noninvasive population measures of neural activity such as fMRI and EEG/MEG used in humans and those single-neuron or neuronal population measures employed for studying the visual system in experimental animals.

REFERENCES

Bahrami, B., Lavie, N., & Rees, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Curr. Biol.*, *17*(6), 509–513.

Barnes, J., Howard, R. J., Senior, C., Brammer, M., Bullmore, E. T., Simmons, A., et al. (1999). The functional anatomy of the McCollough contingent colour after-effect. *NeuroReport*, *10*(1), 195–199.

Beck, D. M., Muggleton, N., Walsh, V., & Lavie, N. (2006). Right parietal cortex plays a critical role in change blindness. *Cereb. Cortex*, *16*(5), 712–717.

Beck, D. M., Rees, G., Frith, C. D., & Lavie, N. (2001). Neural correlates of change detection and change blindness. *Nat. Neurosci.*, *4*(6), 645–650.

Block, N. (2007). Consciousness, accessibility and the mesh between psychology and neuroscience. *Behav. Brain Sci.*, *30*(5–6), 481–499.

Bridge, H., Thomas, O., Jbabdi, S., & Cowey, A. (2008). Changes in connectivity after visual cortical brain damage underlie altered visual function. *Brain*, *131*(Pt.6), 1433–1444.

Carmel, D., Lavie, N., & Rees, G. (2006). Conscious awareness of flicker in humans involves frontal and parietal cortex. *Curr. Biol.*, *16*(9), 907–911.

Cowey, A., & Stoerig, P. (1997). Visual detection in monkeys with blindsight. *Neuropsychologia*, *35*(7), 929–939.

Cowey, A., & Walsh, V. (2000). Magnetically induced phosphenes in sighted, blind and blindsighted observers. *NeuroReport*, *11*(14), 3269–3273.

Crick, F., & Koch, C. (2000). The unconscious homunculus. In T. Metzinger (Ed.), *The neuronal correlates of consciousness* (pp. 103–110). Cambridge, MA: MIT Press.

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA*, *95*(24), 14529–14534.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.*, *4*(7), 752–758.

Diaz, M. T., & McCarthy, G. (2007). Unconscious word processing engages a distributed network of brain regions. *J. Cogn. Neurosci.*, *19*(11), 1768–1775.

Driver, J., & Mattingley, J. B. (1998). Parietal neglect and visual awareness. *Nat. Neurosci.*, *1*(1), 17–22.

Driver, J., Vuilleumier, P., Eimer, M., & Rees, G. (2001). Functional magnetic resonance imaging and evoked potential correlates of conscious and unconscious vision in parietal extinction patients. *NeuroImage*, *14*(1, Pt.2), S68–75.

Edelman, D. B., Baars, B. J., & Seth, A. K. (2005). Identifying hallmarks of consciousness in non-mammalian species. *Conscious. Cogn.*, *14*(1), 169–187.

Eriksson, J., Larsson, A., Riklund Ahlstrom, K., & Nyberg, L. (2004). Visual consciousness: Dissociating the neural correlates of perceptual transitions from sustained perception with fMRI. *Conscious. Cogn.*, *13*(1), 61–72.

Evans, S., & Azzopardi, P. (2007). Evaluation of a "bias-free" measure of awareness. *Spat. Vis.*, *20*(1–2), 61–77.

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nat. Neurosci.*, *8*(10), 1380–1385.

Ffytche, D. H., Howard, R. J., Brammer, M. J., David, A., Woodruff, P., & Williams, S. (1998). The anatomy of conscious vision: An fMRI study of visual hallucinations. *Nat. Neurosci.*, *1*(8), 738–742.

Frith, C., Perry, R., & Lumer, E. (1999). The neural correlates of conscious experience: An experimental framework. *Trends Cogn. Sci.*, *3*(3), 105–114.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.*, *3*(8), 837–843.

Hannula, D. E., Simons, D. J., & Cohen, N. J. (2005). Imaging implicit perception: Promise and pitfalls. *Nat. Rev. Neurosci.*, *6*(3), 247–255.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430.

Haynes, J. D., Deichmann, R., & Rees, G. (2005). Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature*, *438*(7067), 496–499.

Haynes, J. D., Driver, J., & Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron*, *46*(5), 811–821.

Haynes, J. D., Lotto, R. B., & Rees, G. (2004). Responses of human visual cortex to uniform surfaces. *Proc. Natl. Acad. Sci. USA*, *101*(12), 4286–4291.

Haynes, J. D., & Rees, G. (2005a). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.*, *8*(5), 686–691.

Haynes, J. D., & Rees, G. (2005b). Predicting the stream of consciousness from activity in human visual cortex. *Curr. Biol.*, *15*(14), 1301–1307.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.*, *7*(7), 523–534.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*(6598), 334–337.

He, S., Cohen, E. R., & Hu, X. (1998). Close correlation between activity in brain area MT/V5 and the perception of a visual motion aftereffect. *Curr. Biol.*, *8*(22), 1215–1218.

He, S., & MacLeod, D. I. (2001). Orientation-selective adaptation and tilt aftereffect from invisible patterns. *Nature*, *411*(6836), 473–476.

Henson, R. N., Mouchlianitis, E., Matthews, W. J., & Kouider, S. (2008). Electrophysiological correlates of masked face priming. *NeuroImage*, *40*(2), 884–895.

Hirsch, J., DeLaPaz, R. L., Relkin, N. R., Victor, J., Kim, K., Li, T., et al. (1995). Illusory contours activate specific regions in human visual cortex: Evidence from functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA*, *92*(14), 6469–6473.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.*, *8*(5), 679–685.

Kamitani, Y., & Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.*, *16*(11), 1096–1102.

Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (2004). Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia*, *42*(6), 831–835.

Kiefer, M., & Brendel, D. (2006). Attentional modulation of unconscious "automatic" processes: Evidence from event-related potentials in a masked priming paradigm. *J. Cogn. Neurosci.*, *18*(2), 184–198.

Kleinschmidt, A., Buchel, C., Zeki, S., & Frackowiak, R. S. (1998). Human brain activity during spontaneously reversing perception of ambiguous figures. *Proc. Biol. Sci.*, *265*(1413), 2427–2433.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends Cogn. Sci.*, *11*(1), 16–22.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *362*(1481), 857–875.

Kouider, S., Dehaene, S., Jobert, A., & Le Bihan, D. (2007). Cerebral bases of subliminal and supraliminal priming during reading. *Cereb. Cortex*, *17*(9), 2019–2029.

Kreiman, G., Koch, C., & Fried, I. (2000). Imagery neurons in the human brain. *Nature*, *408*(6810), 357–361.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.*, *23*(11), 571–579.

Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *J. Neurosci.*, *27*(21), 5805–5811.

Lee, H. W., Hong, S. B., Seo, D. W., Tae, W. S., & Hong, S. C. (2000). Mapping of functional organization in human visual cortex: Electrical cortical stimulation. *Neurology*, *54*(4), 849–854.

Lee, S. H., & Blake, R. (2002). V1 activity is reduced during binocular rivalry. *J. Vis.*, *2*(9), 618–626.

Lee, S. H., Blake, R., & Heeger, D. J. (2005). Traveling waves of activity in primary visual cortex during binocular rivalry. *Nat. Neurosci.*, *8*(1), 22–23.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616–618.

Lumer, E. D., Friston, K. J., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science*, *280*(5371), 1930–1934.

Lumer, E. D., & Rees, G. (1999). Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proc. Natl. Acad. Sci. USA*, *96*(4), 1669–1673.

Marois, R., Yi, D. J., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron*, *41*(3), 465–472.

Meng, M., Remus, D. A., & Tong, F. (2005). Filling-in of visual phantoms in the human brain. *Nat. Neurosci.*, *8*(9), 1248–1254.

Morris, J. S., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc. Natl. Acad. Sci. USA*, *96*(4), 1680–1685.

Moutoussis, K., & Zeki, S. (2002). The relationship between cortical activation and perception investigated with invisible stimuli. *Proc. Natl. Acad. Sci. USA*, *99*(14), 9527–9532.

Moutoussis, K., & Zeki, S. (2006). Seeing invisible motion: A human fMRI study. *Curr. Biol.*, *16*(6), 574–579.

Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biol.*, *3*(8), e265.

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nat. Neurosci.*, *9*(3), 429–434.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.*, *10*(9), 424–430.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.*, *12*(6), 1013–1023.

Overgaard, M., Rote, J., Mouridsen, K., & Ramsoy, T. Z. (2006). Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Conscious. Cogn.*, *15*(4), 700–708.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*(5516), 510–512.

Persaud, N., & Cowey, A. (2008). Blindsight is unlike normal conscious vision: Evidence from an exclusion task. *Conscious. Cogn.*, *17*(3), 1050–1055.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nat. Neurosci.*, *10*(2), 257–261.

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Curr. Opin. Neurobiol.*, *15*(2), 188–196.

Pessoa, L., Japee, S., & Ungerleider, L. G. (2005). Visual awareness and the detection of fearful faces. *Emotion*, *5*(2), 243–247.

Pins, D., & Ffytche, D. (2003). The neural correlates of conscious vision. *Cereb. Cortex*, *13*(5), 461–474.

Polonsky, A., Blake, R., Braun, J., & Heeger, D. J. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nat. Neurosci.*, *3*(11), 1153–1159.

Portas, C. M., Strange, B. A., Friston, K. J., Dolan, R. J., & Frith, C. D. (2000). How does the brain sustain a visual percept? *Proc. Biol. Sci.*, *267*(1446), 845–850.

Rees, G., Russell, C., Frith, C. D., & Driver, J. (1999). Inattentional blindness versus inattentional amnesia for fixated but ignored words. *Science*, *286*(5449), 2504–2507.

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C., & Driver, J. (2000). Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain*, *123*(Pt.8), 1624–1633.

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C., & Driver, J. (2002). Neural correlates of conscious and unconscious vision in parietal extinction. *Neurocase*, *8*(5), 387–393.

Ress, D., & Heeger, D. J. (2003). Neuronal correlates of perception in early visual cortex. *Nat. Neurosci.*, *6*(4), 414–420.

Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Curr. Biol.*, *13*(12), 1038–1041.

Sahraie, A., Trevethan, C. T., MacLeod, M. J., Murray, A. D., Olson, J. A., & Weiskrantz, L. (2006). Increased sensitivity after repeated stimulation of residual spatial channels in blindsight. *Proc. Natl. Acad. Sci. USA*, *103*(40), 14971–14976.

Sahraie, A., Weiskrantz, L., Barbur, J. L., Simmons, A., Williams, S. C., & Brammer, M. J. (1997). Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. *Proc. Natl. Acad. Sci. USA*, *94*(17), 9406–9411.

Sakai, K., Watanabe, E., Onodera, Y., Uchida, I., Kato, H., Yamamoto, E., et al. (1995). Functional mapping of the human colour centre with echo-planar magnetic resonance imaging. *Proc. Biol. Sci.*, *261*(1360), 89–98.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.*, *8*(10), 1391–1400.

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol. Sci.*, *15*(11), 720–728.

Seth, A. K. (2008). Post-decision wagering measures metacognitive content, not sensory consciousness. *Conscious. Cogn.*, *17*(3), 981–983.

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: An MEG study. *Neurosci. Lett.*, *378*(2), 76–81.

Silbersweig, D. A., Stern, E., Frith, C., Cahill, C., Holmes, A., Grootoonk, S., et al. (1995). A functional neuroanatomy of hallucinations in schizophrenia. *Nature*, *378*(6553), 176–179.

Silvanto, J., Cowey, A., Lavie, N., & Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. *Nat. Neurosci.*, *8*(2), 143–144.

Srinivasan, R., Russell, D. P., Edelman, G. M., & Tononi, G. (1999). Increased synchronization of neuromagnetic responses during conscious perception. *J. Neurosci.*, *19*(13), 5435–5448.

Sterzer, P., Haynes, J.-D., & Rees, G. (2008). Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. *J. Vis.*, *8*(15), 10, 1–12.

Sterzer, P., Haynes, J. D., & Rees, G. (2006). Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5. *NeuroImage*, *32*(3), 1308–1316.

Sterzer, P., & Rees, G. (2008). A neural basis for percept stabilization in binocular rivalry. *J. Cogn. Neurosci.*, *20*(3), 389–399.

Sterzer, P., Russ, M. O., Preibisch, C., & Kleinschmidt, A. (2002). Neural correlates of spontaneous direction reversals in ambiguous apparent visual motion. *NeuroImage*, *15*(4), 908–916.

Struber, D., & Herrmann, C. S. (2002). MEG alpha activity decrease reflects destabilization of multistable percepts. *Brain Res. Cogn. Brain Res.*, *14*(3), 370–382.

Sumner, P., Tsai, P. C., Yu, K., & Nachev, P. (2006). Attentional modulation of sensorimotor processes in the absence of perceptual awareness. *Proc. Natl. Acad. Sci. USA*, *103*(27), 10520–10525.

Super, H., Spekreijse, H., & Lamme, V. A. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nat. Neurosci.*, *4*(3), 304–310.

Tong, F., & Engel, S. A. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature*, *411*(6834), 195–199.

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends Cogn. Sci., (10)*11, 502–511.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*(4), 753–759.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, *282*(5395), 1846–1851.

Tononi, G., Srinivasan, R., Russell, D. P., & Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proc. Natl. Acad. Sci. USA*, *95*(6), 3198–3203.

Tootell, R. B., Reppas, J. B., Dale, A. M., Look, R. B., Sereno, M. I., Malach, R., et al. (1995). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature*, *375*(6527), 139–141.

Turatto, M., Sandrini, M., & Miniussi, C. (2004). The role of the right dorsolateral prefrontal cortex in visual change awareness. *NeuroReport*, *15*(16), 2549–2552.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.*, *24*(6), 1656–1674.

Vuilleumier, P., Sagiv, N., Hazeltine, E., Poldrack, R. A., Swick, D., Rafal, R. D., et al. (2001). Neural fate of seen and unseen faces in visuospatial neglect: A combined event-related functional MRI and event-related potential study. *Proc. Natl. Acad. Sci. USA*, *98*(6), 3495–3500.

Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, *31*(3), 1247–1256.

Weiskrantz, L., Rao, A., Hodinott-Hill, I., Nobre, A. C., & Cowey, A. (2003). Brain potentials associated with conscious aftereffects induced by unseen stimuli in a blindsight subject. *Proc. Natl. Acad. Sci. USA*, *100*(18), 10500–10505.

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *J. Neurosci.*, *24*(12), 2898–2904.

Wunderlich, K., Schneider, K. A., & Kastner, S. (2005). Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat. Neurosci.*, *8*(11), 1595–1602.

Wyart, V., & Tallon-Baudry, C. (2008). Neural dissociation between visual awareness and spatial attention. *J. Neurosci.*, *28*(10), 2667–2679.

Zeki, S. (2008). The disunity of consciousness. *Prog. Brain Res.*, *168*, 11–18.

Zeki, S., Watson, J. D., & Frackowiak, R. S. (1993). Going beyond the information given: The relation of illusory visual motion to brain activity. *Proc. Biol. Sci.*, *252*(1335), 215–222.

81 The Role of Feedback in Visual Attention and Awareness

STEPHEN L. MACKNIK AND SUSANA MARTINEZ-CONDE

ABSTRACT The mammalian visual system includes numerous brain areas that are profusely interconnected. With few exceptions, these connections are reciprocal. Anatomical feedback connections in general outnumber feedforward connections, leading to widespread speculation that feedback connections play a critical role in visual awareness. However, evidence from physiological experiments suggests that feedback plays a modulatory role, rather than a driving role. Here we discuss theoretical constraints on the significance of feedback's anatomical numerical advantage, and we describe theoretical limits on feedback's potential physiological impact. These restrictions confine the potential role of feedback in visual awareness and rule out some extant models of visual awareness that require a fundamental role of feedback. We propose that the central role of feedback is to maintain visuospatial attention, rather than visual awareness. Our conclusions highlight the critical need for experiments and models of visual awareness that control for the effects of attention. As a matter of clarity in this chapter: by "visual awareness" or "visibility" we mean the conscious perception that a stimulus is visible. Thus, for the purposes of this discussion, we use the terms *visual awareness*, *visibility*, and *consciousness* interchangeably.

Anatomical observations of feedback in the visual system

The visual areas of the brain are interconnected in a complex pattern of feedforward, lateral, and feedback pathways (Felleman & Van Essen, 1991). Feedback connections are ubiquitous throughout the cortex, and subcortical regions in ascending hierarchical pathways also receive a large amount of feedback from cortical areas (Erisir, Van Horn, & Sherman, 1997; Fitzpatrick, Usrey, Schofield, & Einstein, 1994; Guillery, 1969; Sherman & Guillery, 2002).

ANATOMY OF FEEDBACK IN THE LGN Corticogeniculate input is the largest source of synaptic afferents to the cat lateral geniculate nucleus (LGN). Whereas retinal afferents only encompass 25% of the total number of inputs to LGN interneurons, 37% of the synaptic contacts come from the cortex. In the case of relay cells, the respective percentages are 12% versus 58% (Montero, 1991). Similar estimates have been calculated in the primate, and the general agreement is that the cortical-to-retinal input ratio is between 1:2 and 1:6 in both cats and primates (Erisir et al., 1997; Fitzpatrick et al., 1994; Guillery, 1969; Sherman & Guillery, 2002; Van Horn, Erisir, & Sherman, 2000). Boyapati and Henry (1984) concluded that feedback connections from the cat visual cortex to the LGN concentrated a larger fraction of fine axons than feedforward geniculocortical connections, presumably resulting in comparatively slower conduction speeds.

These and other considerations concerning the synaptic size, efficacy, and contribution of feedback connections underscore the potential mistake in assuming that a numerically larger number of inputs means that those inputs are functionally most important (Sherman & Guillery, 2002).

ANATOMY OF FEEDBACK IN THE PRIMARY VISUAL CORTEX Cortical feedforward pathways usually project from the supragranular layers of visual areas early in the hierarchy (less than 10–15% of the connections may arise from deep layers) and terminate in layer 4 of areas later in the hierarchy. In contrast, feedback projections usually arise from the infragranular layers of later areas and terminate outside of layer 4 in the early areas (Barone, Batardiere, Knoblauch, & Kennedy, 2000; Felleman & Van Essen, 1991; Hilgetag, O'Neill, & Young, 1996a, 1996b; Maunsell & Van Essen, 1983).

Direct feedforward projections to primate area V1 (also called primary visual cortex, striate cortex, and Brodmann's area 17) originate from the pulvinar, LGN, claustrum, nucleus paracentralis, raphe system, locus coeruleus, and nucleus basalis of Meynert (Blasdel & Lund, 1983; Doty, 1983; Fitzpatrick et al., 1994; Hendry & Yoshioka, 1994; Lachica & Casagrande, 1992; Ogren & Hendrickson, 1976; Perkel, Bullier, & Kennedy, 1986; Rezak & Benevento, 1979). Direct feedforward projections from V1 extend to V2, V3, V5 or MT, MST, and FEF (Boussaoud, Ungerleider, & Desimone, 1990; Livingstone & Hubel, 1987; Lund, Lund, Hendrickson, Bunt, & Fuchs, 1975; Maunsell & Van Essen, 1983; Shipp & Zeki, 1989; Ungerleider & Desimone, 1986a, 1986b). Direct feedback projections to V1 originate from V2, V3, V4, V5 or MT, MST, FEF, LIP, and inferotemporal cortex (Barone et al., 2000; Perkel et al., 1986; Rockland, Saleem, & Tanaka, 1994; Shipp & Zeki,

STEPHEN L. MACKNIK and SUSANA MARTINEZ-CONDE Barrow Neurological Institute, Phoenix, Arizona

1989; Suzuki, Saleem, & Tanaka, 2000; Ungerleider & Desimone, 1986a, 1986b). Direct feedback projections from V1 extend to SC, LGN, pulvinar, and pons (Fitzpatrick et al., 1994; Fries, 1990; Fries & Distel, 1983; Gutierrez & Cusick, 1997; Lund et al., 1975).

Peters, Payne, and Budd (1994) showed that only 1–8% of the synaptic inputs into a layer 4C neuron from primate area V1 originate in the LGN. They concluded that "it is unlikely that the response properties of a particular cortical neuron are dominated by its input from a single geniculate neuron" (p. 215). However, this conclusion was based solely on the anatomical numbers of inputs, and not on their functional properties, which we discuss further in the next section.

Physiological observations of feedback in the visual system

Methodological Shortcomings in Physiological Studies of Feedback Some visual physiology studies have found that feedback connections between the secondary and primary visual cortices enhance or decrease neuronal responsiveness without fundamentally altering response specificity (Martinez-Conde et al., 1999) (see figure 81.1). These studies were conducted by microinjecting small amounts of neuronal modulators into area 18 of the cat while recording from the corresponding retinotopic position in area 17. This method is accurate in its assessment of feedback effects because it sequesters the source of neuronal enhancement and suppression to a small focal region that cannot directly affect the neuronal responses of the neurons being recorded in the area of interest. Thus the only possible cause of the response modulation in area 17 was the feedback connection from area 18. Another positive aspect of this technique is that the effects are fully reversible, which is not a feature shared by the ablation (Super & Lamme, 2007) and lesion methods.

Other studies have proposed a more significant physiological role for feedback in the visual system. However, these studies have generally used alternative methods such as ablation, cooling, transcranial magnetic stimulation, and direct pharmacological manipulation of the neurons being recorded. Such techniques are usually disadvantageous in that they are nonfocal, nonreversible, and/or may have unknown or poorly understood nonspecific effects on the physiological milieu of the neurons being directly recorded (such as by changing the pH, osmolarity, temperature, or other effects). Nonfocal and/or nonreversible techniques may also affect the vasculature feeding the targeted neurons or fibers of passage with known or unknown connectivity (either direct or indirect) to the targeted neurons. Thus the results obtained are more difficult to interpret, as the responses of the targeted neurons may have been affected in ways unrelated to any putative role of feedback.

Also, as we will discuss more fully in a later section, it is critical that physiological measurements of feedback, as they relate to awareness, be conducted with careful controls for the effects of attention, as well as its underlying circuits. This is a necessary precaution, as the physiological process of attention is differentiable from that of awareness (Koch & Tsuchiya, 2007). For a comprehensive discussion of this issue, please see Koch (chapter 79, this volume).

What Do We Mean by Feedback? For the purposes of this chapter, we restrict our definition of the word "feedback" to the long-range fibers that connect a higher brain area to a lower brain area, within an ascending sensory system. In this definition, the same information arrives to the same neural circuit at least twice: first as it feeds forward through the system and later again as it feeds back.

Other types of feedback loops in the brain are not discussed in this chapter. For instance, information may flow up from one thalamic nucleus to the cortex (i.e., from the LGN to area V1) and then back down to a different thalamic nucleus (i.e., the pulvinar) (Rockland, 1996). In this case, it could be said that the thalamus as a whole sends information to, and receives feedback from, area V1 (as Rockland describes it). However, the feedforward and feedback projections are mediated by two separate thalamic nuclei. Thus this type of circuit does not meet the definition of feedback used here.

Here we will discuss specifically those reciprocal connections between visual areas of the geniculate-cortical pathway. Thus the feedback we will consider entails connections from neurons processing more complex visual information (and having more complex receptive fields) to neurons processing less complex visual information (which have simpler and less selective receptive fields) (figure 81.2).

This chapter aims to describe the powerful constraints on the functional role of feedback, even within such an ordinary and basic neural system. Figure 81.2*A* illustrates the basic connectivity between an area of the geniculocortical pathway and the next area up in the hierarchy (i.e., the LGN and area V1). The lower, simpler level of processing feeds forward to the higher level. There, information is further processed by neural circuits with more complex receptive fields. The higher, more complex level then feeds back information to the simpler level. If such a feedback connection is functionally effective, the receptive fields from the lower level will acquire the specificity and complexity that characterize the higher-level receptive fields (figure 81.2*B*). This physiological prediction should apply to any feedback pathways that are both engaged and significant in strength.

Physiology of Feedback in the LGN Corticogeniculate connections to the LGN are retinotopically organized, and

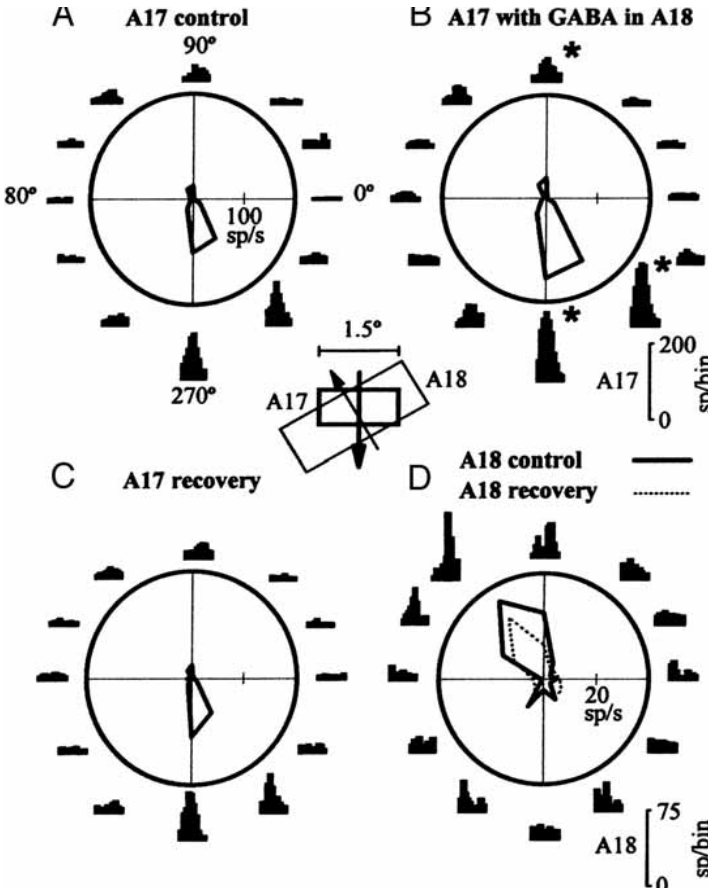


FIGURE 81.1 Reversible removal of feedback from area 18 to area 17 in the cat. (*A*) Orientation tuning curve of a cell from layers 2/3 of area 17. (*B*) Orientation tuning curve of the same cell during GABA application in area 18. Note that, although the firing rate of this area 17 neuron increases significantly, its orientation selectivity is virtually unchanged in the absence of feedback from area 18. Thus feedback modulates the magnitude of the neuronal responses but does not affect their functional specificity. (*C*) Orientation tuning curve of the same cell after area 18 blockade. (*D*) Solid line, control tuning curve of an area 18 cell recorded simultaneously. Dotted line, tuning curve of the same cell after blockade. *Inset*: receptive fields of both cells. For clarity, only the preblockade post-stimulus time histograms (PSTHs) are shown in *D*. The number of spikes for PSTHs of areas 17 and 18 are indicated at *bottom* of *B* and *D*, respectively. Bin size: 100 ms. Time base: 1 s. (Reprinted from Martinez-Conde et al., 1999.)

they preferentially end on LGN layers with the same ocular dominance as the cortical cells of origin (Murphy & Sillito, 1996). Although only 12–58% (Montero, 1991) of the inputs to geniculate cells are retinal in origin, these synapses drive the primary responses of geniculate relay cells, whereas feedback inputs play a modulatory role (Sherman & Guillery, 1998, 2002).

Physiology of Feedback in the Primary Visual Cortex

Cortico-cortical feedback connections are also retinotopically specific (Salin, Girard, Kennedy, & Bullier, 1992). For instance, there is a functional projection from area 18 to area 17 neurons with similar retinotopic locations (Bullier, McCourt, & Henry, 1988; Martinez-Conde et al., 1999; Salin et al., 1992; Salin, Kennedy, & Bullier, 1995). Girard, Hupe, and Bullier (2001) found that feedforward and feedback connections between areas V1 and V2 of the monkey have similarly rapid conduction speeds.

In the cat visual cortex, electrical stimulation from areas area 18 and area 19 demonstrated 50% of monosynaptic connections with superficial layers of area 17, in regions with similar functional properties, such as retinotopic location (Bullier et al., 1988). Mignard and Malpeli (1991) also found that inactivation of area 18 in the cat led to decreased responses in area 17. Martinez-Conde and colleagues (1999) found that focal reversible inactivation of area 18 produced suppressed or enhanced visual responses in area 17 neurons with a similar retinotopy. In most area 17 neurons, orientation bandwidths and other functional characteristics remained unaltered, suggesting that feedback from area 18 modulates area 17 responses without fundamentally altering their specificity.

In the squirrel monkey, Sandell and Schiller (1982) found that most area V1 cells decreased their visual responses when area V2 was reversibly cooled, although a few cells became more active. Orientation selectivity remained

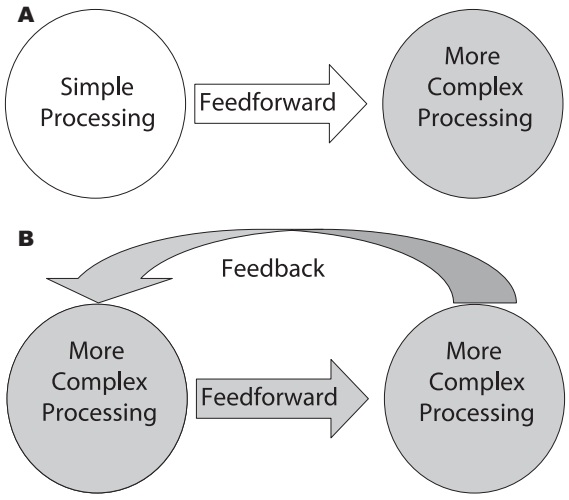


FIGURE 81.2 A generalized model of the effect of feedback in a hierarchy of simple to complex neural processing. (*A*) In a functional hierarchy, information processing becomes more complex as one ascends in the pathway. (*B*) When feedback is engaged, the lower levels of the hierarchy take on the more complex properties of the upper levels.

unchanged, although direction selectivity decreased in some instances. Bullier, Hupe, James, and Girard (1996) reported in the cynomologous monkey that, following GABA inactivation of area V2, V1 neurons showed decreased or unchanged responses in the center of the classical receptive field, but increased responses in the region surrounding it. These results were supported by subsequent findings in areas V1, V2, and V3 following area MT inactivation (Hupe et al., 1998). More recently, Angelucci and colleagues (Angelucci & Bressloff, 2006; Angelucci, Levitt, & Lund, 2002) have suggested that area V1 extraclassical receptive field properties arise from area V2 feedback.

In summary, physiological studies as a whole suggest that feedback connections in the visual system may play a modulatory role, rather than a driving role, in shaping the responses of hierarchically lower areas. This evidence agrees with the "no-strong-loops" hypothesis formulated by Crick and Koch (1998b). The no-strong-loops hypothesis proposes that all strong connections in the visual system are of the feedforward type. That is, "the visual cortex is basically a feedforward system that is modulated by feedback connections," which is "not to say that such modulation may not be very important for many of its functions" (p. 248). Crick and Koch argued that "although neural nets can be constructed with feedback connections that form loops, they do not work satisfactorily if the excitatory feedback is too strong." Similarly, if feedback connections formed "strong, directed loops" in the brain, the cortex would as a result "go into uncontrolled oscillations." Therefore, the relative number of feedback versus feedforward anatomical connections to any given visual area may be misleading as to the respective roles of such connections. For instance, the fact that the cat LGN receives substantially larger numbers of synapses from the cortex than from the retina (Montero, 1991) does not necessarily mean that corticogeniculate connections are more important than retinogeniculate connections in determining the response characteristics of LGN neurons.

Although the role of feedback modulation in our visual perception remains unclear, one possibility is that feedback may be involved in attentional mechanisms (Martinez-Conde et al., 1999). We will discuss this idea more fully in the next section.

The role of feedback in attention

Based on the evidence we have reviewed, one potentially important role for feedback may be to carry attentional modulation signals. Other modulatory roles for feedback remain possible, but none are as clearly established. Thus it may be that the sole effect of all feedback connectivity is to facilitate and suppress attention. At first, given the massive amount of anatomical feedback versus feedforward connections, this possibility may seem unlikely. Indeed, the great extent of feedback connectivity suggests to some that feedback must have a large number of roles (Sherman & Guillery, 2002; Sillito & Jones, 1996). However, we will argue here that the great number of feedback connections may potentially be explained by the need for top-down attentional modulation alone. Ascending circuits in the visual system primarily form a labeled-line hierarchy, and so feedback connections necessarily require more wiring than feedforward connections to send back even the simplest signal.

To illustrate the logic of this argument, let us consider the anatomical connectivity between the LGN and V1 (figure 81.3). As previously described, LGN relay cells receive more numerous feedback connections from the cortex than feedforward inputs from the retina. However, because V1 receptive fields are orientation selective and LGN receptive fields are not, any functionally significant feedback from V1 to a given retinotopic location in the LGN must represent many, or all, orientations. (One should note that Vidyasagar & Urbas, 1982, found slight orientation biases in LGN receptive fields; these biases were much smaller than the strong orientation selectivity found in V1.) That is, for each unoriented feedforward connection from the LGN to V1, there *must be* many oriented feedback connections from V1 to the LGN, each with a different orientation, so that the sum of all feedback projections spans all the orientation space. Otherwise, if the orientation space of the feedback connections were not filled completely, LGN receptive fields would show a substantial orientation bias. Thus anatomical feedback connectivity must be large so as to represent the entire orientation space at each retinotopic location. However, because

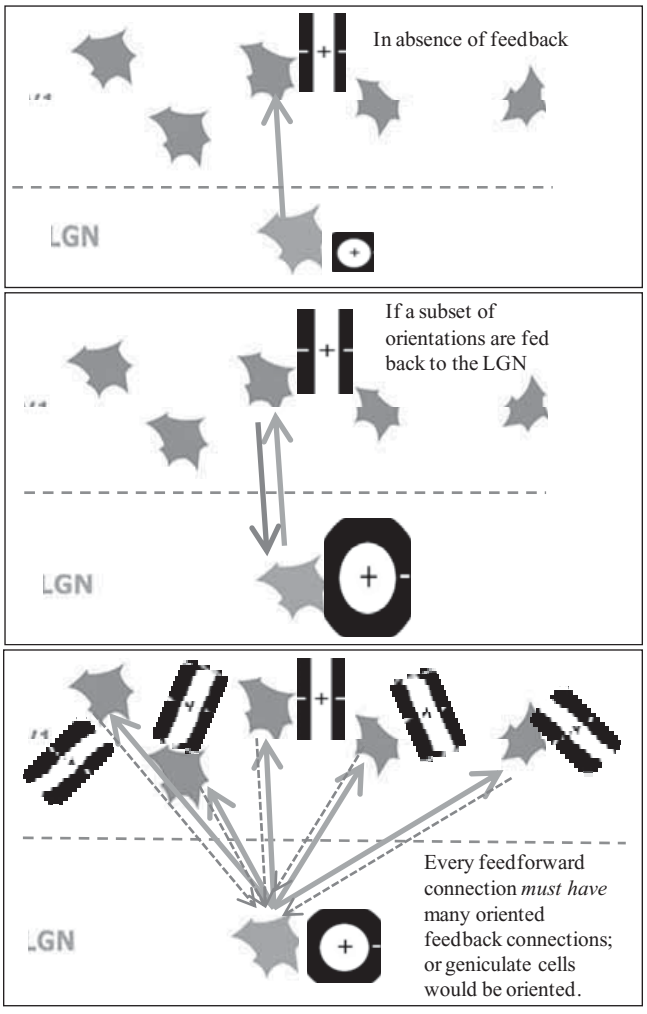


FIGURE 81.3 A model of the effects of feedback from area V1 on an LGN neuron. Numerous V1-oriented cells must feed back to every feedforward LGN cell in order to account for the lack of significant orientation bias in LGN receptive fields.

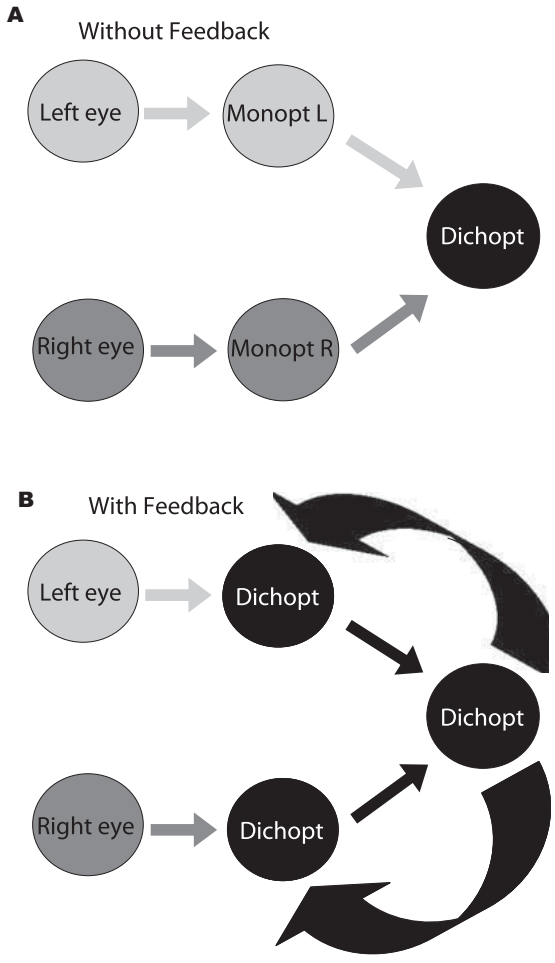


FIGURE 81.4 The effects of feedback from dichoptic levels to monoptic levels of visual processing. (*A*) A general model of early visual binocular integration in the absence of feedback connections. (*B*) If significant feedback existed between dichoptic levels of processing and earlier monoptic levels, the earlier levels should acquire the properties of the dichoptic levels (i.e., they should become dichoptic by virtue of the feedback). (Reprinted from Macknik, 2006.)

of their orientation selectivity, only a fraction of all feedback connections will be active at any given time, depending on the orientation of the visual stimulus, whereas the feedforward connections will be active irrespective of stimulus orientation. In summary, the massive feedback versus feedforward connectivity ratio can be misleading: this large ratio does not necessarily mean that feedback signals are more important or more physiologically relevant than feedforward signals. Because higher visual areas are more selective than lower visual areas, only a relatively small fraction of the feedback may be expected to be active at any given moment. Thus feedback connections may need to tile the entire space of receptive field properties of the higher level; otherwise, the feedback would impose high-level receptive field characteristics on the receptive fields of lower areas.

Figure 81.4 illustrates this idea in terms of the feedback from dichoptic to monoptic levels of the visual pathway. To be clear about the jargon: "monocular" means "with respect to a single eye," and "monoptic" means either "monocular" or "not different between the two eyes." "Binocular" means "with respect to both eyes," and "dichoptic" means "different in the two eyes." Thus subcortical levels of the visual system are monoptic (because the cells are monocular), whereas cortical visual areas that have binocular circuits may potentially process dichoptic information.

To summarize: because receptive fields in ascending pathways become more selective, larger, and more complex in their properties as one rises through the higher levels of the brain's hierarchies, anatomical feedback connections must be more numerous than feedforward connections. Otherwise, the hierarchical nature of the visual system would be diminished (figure 81.2). Moreover, the numerical

advantage of feedback over feedforward connections should be expected even if there is just a single functional role for feedback (i.e., attentional modulation).

If we combine these ideas with the Crick and Koch's no-strong-loops hypothesis and the physiological findings indicating that feedback plays a modulatory rather than a driving role, we may conclude that feedback inputs have more moderate physiological effects than feedforward inputs, despite being anatomically more numerous. This concept is supported by the known physiology: besides their lack of orientation selectivity, another feature that distinguishes LGN from V1 receptive fields is their smaller size (Allman, Miezin, & McGuinness, 1985; Desimone, Schein, Moran, & Ungerleider, 1985; Kastner, Nothdurft, & Pigarev, 1999; Knierim & Van Essen, 1991; Zeki, 1978a, 1978b). If feedback connections from V1 to the LGN were functionally as strong as their feedforward counterparts, LGN receptive fields would be as large as V1 receptive fields, but they are not. That is, because LGN receptive fields are smaller than V1 receptive fields, feedback from V1 to the LGN must be weaker than the retinal inputs.

It follows from these ideas that when feedback is operational, some receptive field properties, such as size, which continues to increase throughout the visual hierarchy (Allman et al., 1985; Desimone et al., 1985; Kastner et al., 1999; Knierim & Van Essen, 1991; Zeki, 1978a, 1978b) will be fed back from higher to lower levels. Thus we may predict that, if attention is carried by feedback connections, earlier receptive fields should increase in size when attention is applied actively. This prediction has been confirmed experimentally (He, Cavanagh, & Intriligator, 1996; Williford & Maunsell, 2006). Chen and colleagues (2008) recently showed that attentional modulation of V1 neurons in the awake monkey is spatially specific: increasing task difficulty enhanced V1 neuronal firing rate at the focus of attention and suppressed it in surrounding regions, in support of Desimone and Duncan's (1995) center-surround model of attention (figure 81.5). Moreover, response enhancement and suppression were mediated by distinct neuronal populations that differed in direction selectivity, spike width, interspike-interval distribution and contrast sensitivity. This finding suggested that attentional feedback facilitates and suppresses distinct populations of neurons in the primary visual cortex.

To conclude, feedback connections may potentially have no other function than to modulate (facilitate or suppress) feedforward signals as a function of attentional load.

The role of visual masking, binocular rivalry, attention, and feedback in the study of visual awareness

Let us assume that visual awareness is correlated to brain activity within specialized neural circuits, and that not all brain circuits maintain awareness. It follows that the neural activity that leads to reflexive or involuntary motor action may not correlate with awareness because it does not reside within awareness-causing neural circuits (Macknik & Martinez-Conde, 2009).

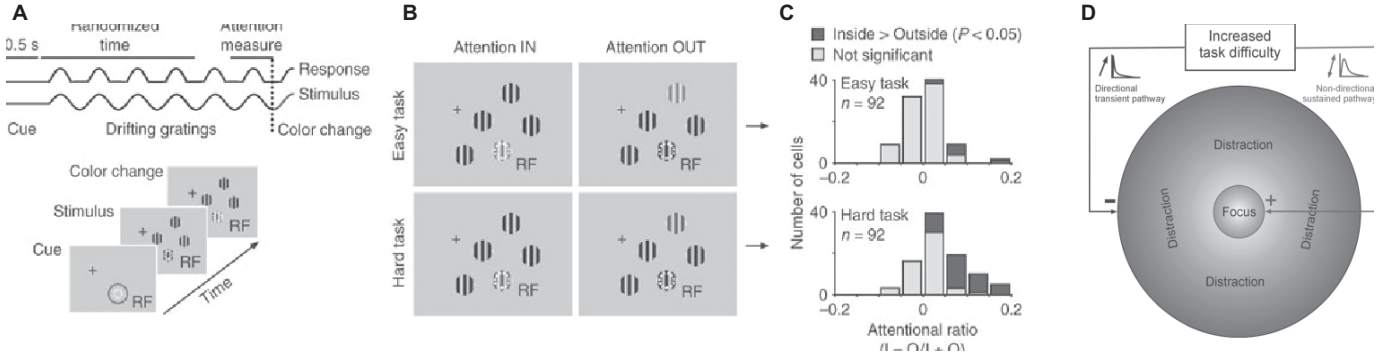


FIGURE 81.5 V1 attentional response modulation in awake monkey single cells during hard and easy tasks. (*A*) Temporal structure of a trial. Two rhesus monkeys were trained to fixate on a small cross while covertly attending to a spatial location that was cued at the beginning of each trial. The cue was a thin red ring with a diameter that was threefold larger than the diameter of the neuronal receptive field (RF). Following the cue, drifting gratings were presented simultaneously at five different spatial locations for 1.5–3 s. Following a randomized period of time, one of the gratings changed color/luminance, and the animal was tasked with detecting the change by releasing a bar within 0.5 s. The attentional modulations were measured at the last cycle of the drifting grating before the color change. (*B*) The color change could be easy or hard to detect and could occur inside or outside of the receptive field. (*C*) The number of cells that were significantly modulated by attention (red) was much lower during the easy task (top) than during the hard task (bottom). A subset of 8 cells was significantly modulated by attention during both the easy and the hard task ($P < 0.05$). (*D*) An increase in task difficulty leads to an enhancement of V1 visual responses at the focus of attention and a suppression outside the focus. Difficulty-enhanced V1 neurons have poor directional selectivity and broad interspike interval distributions (sustained responses). Difficulty-suppressed V1 neurons are directional selective and have tight interspike interval distributions (transient responses). (Reprinted from Chen et al., 2008.)

Let us also propose that there is a "minimal set of neural conditions" necessary to achieve conscious visibility (see Chalmers, 2000, for an excellent review of this idea). Such conditions take the form of a specific type (or types) of neural activity within a subset of brain circuits. The minimal set of conditions will not be met if the correct circuits have the wrong type of activity (too much activity, too little activity, sustained activity when transient activity is required, etc). Moreover, if the correct type of activity occurs, but solely within circuits that do not maintain awareness, visibility will also fail. Finding the conditions in which visibility fails is critical to the research described here; although we do not yet know what the minimal set of conditions is, we can nevertheless systematically modify potentially important conditions (change neural circuits, modify levels of activity) and see if they result in stimulus invisibility. If so, the modified condition would be part, potentially, of the minimal set of neural conditions necessary to maintain visibility.

To establish the minimal set of conditions for visibility we need to answer at least four questions (Macknik, 2006). The questions and their (partial) answers are as follows:

1. What stimulus parameters are important to visibility?
   A. The spatiotemporal edges (also curves and corners) of stimuli are the most important parameters to stimulus visibility (Macknik, Martinez-Conde, & Haglund, 2000; Troncoso, Macknik, & Martinez-Conde, 2005; Troncoso et al., 2007).
2. What types of neural activity best maintain visibility (transient versus sustained firing, rate codes, bursts of spikes, etc.—that is, what is the neural code for visibility)?
   A. Transient bursts of spikes best maintain visibility (Macknik & Livingstone, 1998; Macknik et al., 2000; Martinez-Conde, Macknik, & Hubel, 2000, 2002). See figure 81.6.
3. What brain areas must be active to maintain visibility?
   A. Visual areas downstream of V2, lying within the occipital lobe, must be active to maintain visibility of simple unattended targets (Macknik, 2006; Macknik & Martinez-Conde, 2004a; Tse, Martinez-Conde, Schlegel, & Macknik, 2005).
4. What specific neural circuits within the relevant brain areas maintain visibility?
   A. The specific circuits that maintain visibility are currently unknown, but their responsivity is modulated by lateral inhibition (Macknik, 2006; Macknik & Livingstone, 1998; Macknik & Martinez-Conde, 2004a, 2004b; Macknik et al., 2000).

We must also determine the set of standards that will allow us to conclude that any given brain area, or neural circuit within an area, is responsible for generating a conscious

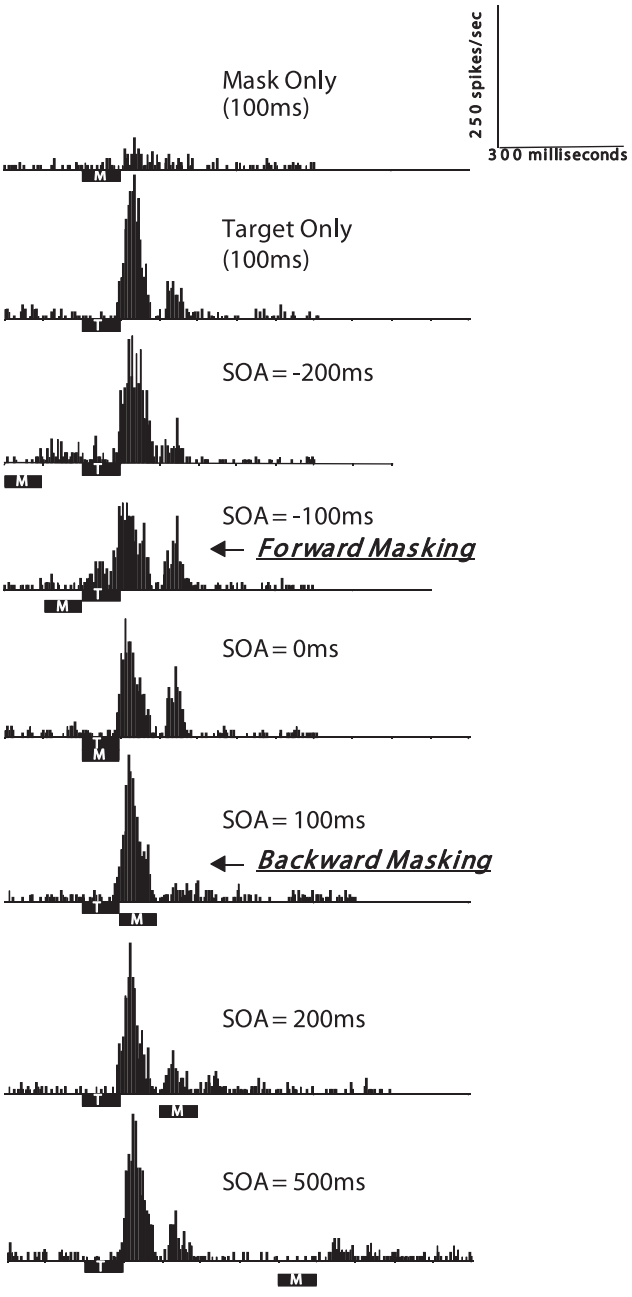


FIGURE 81.6 Multiunit recording from upper layers of area V1 in an anesthetized rhesus monkey. Black boxes below each histogram represent the time course of the mask (*M*) and target (*T*). Notice that under conditions that best correlate with human forward masking (interstimulus interval [ISI] of 0 ms, here corresponding to stimulus onset asynchrony [SOA] of −100 ms), the main effect of the mask is to inhibit the transient onset-response to the target. Similarly, in the condition that produces maximum backward masking in humans (stimulus termination asynchrony [STA] of 100 ms, here corresponding to SOA of 100 ms), the afterdischarge is specifically inhibited. Each histogram is an average of 50 trials with a bin width of 5 ms. (Reprinted from Macknik & Livingstone, 1998.)

experience. Parker and Newsome developed a "list of idealized criteria that should be fulfilled if we are to claim that some neuron or set of neurons plays a critical role in the generation of a perceptual event" (Parker & Newsome, 1998, p. 230). If one replaces the words "perceptual event" with "conscious experience," Parker and Newsome's list can be used as an initial foundation for the neurophysiological requirements needed to establish whether any given neuron or brain circuit may be the neural substrate of awareness (Macknik & Martinez-Conde, 2007). Parker and Newsome's list (pp. 230–231) follows:

1. The responses of the neurons and of the perceiving subject should be measured and analyzed in directly comparable ways.

2. The neurons in question should signal relevant information when the organism is carrying out the chosen perceptual task. Thus the neurons should have discernible features in their firing patterns in response to the different external stimuli that are presented to the observer during the task.

3. Differences in the firing patterns of some set of the candidate neurons to different external stimuli should be sufficiently reliable in a statistical sense to account for, and be reconciled with, the precision of the organism's responses.

4. Fluctuations in the firing of some set of the candidate neurons to the repeated presentation of identical external stimuli should be predictive of the observer's judgment on individual stimulus presentations.

5. Direct interference with the firing patterns of some set of the candidate neurons (e.g., by electrical or chemical stimulation) should lead to some form of measurable change in the perceptual responses of the subject at the moment that the relevant external stimulus is delivered.

6. The firing patterns of the neurons in question should not be affected by the particular form of the motor response that the observer uses to indicate his or her percept.

7. Temporary or permanent removal of all or part of the candidate set of neurons should lead to a measurable perceptual deficit, however slight or transient in nature.

However, visual circuits that may pass muster with Parker and Newsome's guidelines may nevertheless fail to maintain awareness, as we shall explain. To guide the search for the neural correlates of consciousness (NCC), the minimal neuronal mechanisms jointly sufficient for a particular percept (Crick & Koch, 1995), some additional standards must be added.

The first additional standard concerns the use of illusions as the tool of choice to test whether a neuronal population or circuit may maintain awareness. Visual illusions, by definition, dissociate the subject's perception of a stimulus from its physical reality. Thus visual illusions are powerful devices in the search for the NCC (Myerson, Miezin, & Allman, 1981), as they allow us to distinguish the neural responses to the physical stimulus from the neural responses that correlate to perception. Our brains ultimately construct our perceptual experience, rather than reconstruct the physical world (Macknik & Haglund, 1999). Therefore an awareness-maintaining circuit should express activity that matches the conscious percept, irrespective of whether it matches the physical stimulus. Neurons (circuits, brain areas) that produce neural responses that fail to match the percept provide the most useful information because they can be ruled out, unambiguously, as part of the NCC. As a result, the search for the NCC can be focused to the remaining neural circuits. Conversely, neurons that do correlate with perception are not necessarily critical to awareness, as they may simply play a support role (among other possibilities) without causing awareness themselves.

The second additional standard derives from a major contribution of Crick and Koch's: the distinction between explicit and implicit representations in the study of visual awareness (Crick & Koch, 1998a). In an explicit representation of a stimulus feature, there is a set of neurons that represents that feature without substantial further processing. In an implicit representation, the neuronal responses may account for certain elements of a given feature; however, the feature itself is not detected at that level. For instance, all visual information is implicitly encoded in the photoreceptors of the retina. The orientation of a stimulus, however, is not explicitly encoded until area V1, where orientation-selective neurons and functional orientation columns are first found. Crick and Koch propose that there is an explicit representation of every conscious percept.

Here we offer the following corollary to Crick and Koch's idea of explicit representation: Before one can test a neuronal population or circuit for its role in the NCC, the specific neurons (or the population/circuit being tested) must be shown to explicitly process the test stimulus. That is, the neurons must respond to the test stimulus or show selective tuning to some range of features of the stimulus. This corollary constrains the design of neurophysiological experiments aimed to test the participation of specific neurons, circuits, and brain areas in the NCC. For instance, if one found that retinal responses do not correlate with auditory awareness, such a discovery would not carry great weight. The neurons in the eye do not process auditory information, and so it is not appropriate to test their correlation to auditory perception. However, this caveat also applies to more nuanced stimuli. What if V1 activity was tested for its correlation to the perception of faces versus houses? Faces and houses are visual stimuli, but V1 has never been shown to process faces or houses explicitly, despite the fact that visual information about faces and houses must implicitly be represented in V1. Therefore, one cannot test V1's role in the NCC using houses versus faces and expect to come to any meaningful

conclusion. Because that form of information is not explicitly processed in V1, it would not be informative as to V1's role in the NCC if V1 neurons failed to modulate their response when the subject was presented with faces versus houses. It follows that some stimuli are incapable of localizing awareness within specific neural circuits, because no appropriate control exists to test for their explicit representation. For this reason, binocular rivalry stimuli pose a special problem in localizing the circuits that maintain visual awareness. Binocular rivalry (Wheatstone, 1838) is a dynamic percept that occurs when two disparate images that cannot be fused stereoscopically are presented dichoptically to the subject (i.e., each image is presented independently to each of the subject's eyes). The two images (or perhaps the two eyes) appear to compete with each other, and the observer perceives repetitive undulations of the two images, so that only one of them dominates perceptually at any given time. (If the images are large enough, then binocular rivalry can occur in a piecemeal fashion, so that parts of each image are contemporaneously visible.)

Binocular rivalry has been used as a tool to assess the NCC, but it has generated controversy because of conflicting results (Macknik & Martinez-Conde, 2004a; Tse et al., 2005). Some human fMRI studies report that BOLD activity in V1 correlates with awareness of binocular rivalry percepts (Lee, Blake, & Heeger, 2005; Polonsky, Blake, Braun, & Heeger, 2000; Tong & Engel, 2001). In contrast, other human fMRI studies (Lumer, Friston, & Rees, 1998), as well as neuronal recording studies in nonhuman primates (Leopold & Logothetis, 1996), report that activity in area V1 does not correlate with visual awareness of binocular rivalry percepts. One possible reason for this discrepancy is that none of these studies determined that the visual areas tested contained the interocular suppression circuits necessary to mediate binocular rivalry. That is, since binocular rivalry is a process of interocular suppression, the neural circuits underlying the perception of binocular rivalry must be shown to produce interocular suppression—explicitly. Otherwise, it cannot be demonstrated that binocular rivalry is a valid stimulus for testing the NCC in those areas. Thus awareness studies using binocular rivalry are valid only in areas that have been shown to maintain interocular suppression. If binocular rivalry fails to modulate activity within a visual area, one cannot know, by using binocular rivalry alone, if the perceptual modulation failed because awareness is not maintained in that area or because the area does not have circuits that drive interocular suppression. This is more than just a theoretical possibility: as we will describe, we have shown in the human and the macaque monkey that the initial binocular neurons of the early visual system (areas V1 and V2) are binocular for excitation but monocular for inhibition. That is, they fail to process interocular suppression explicitly (Macknik & Martinez-Conde, 2004a; Tse

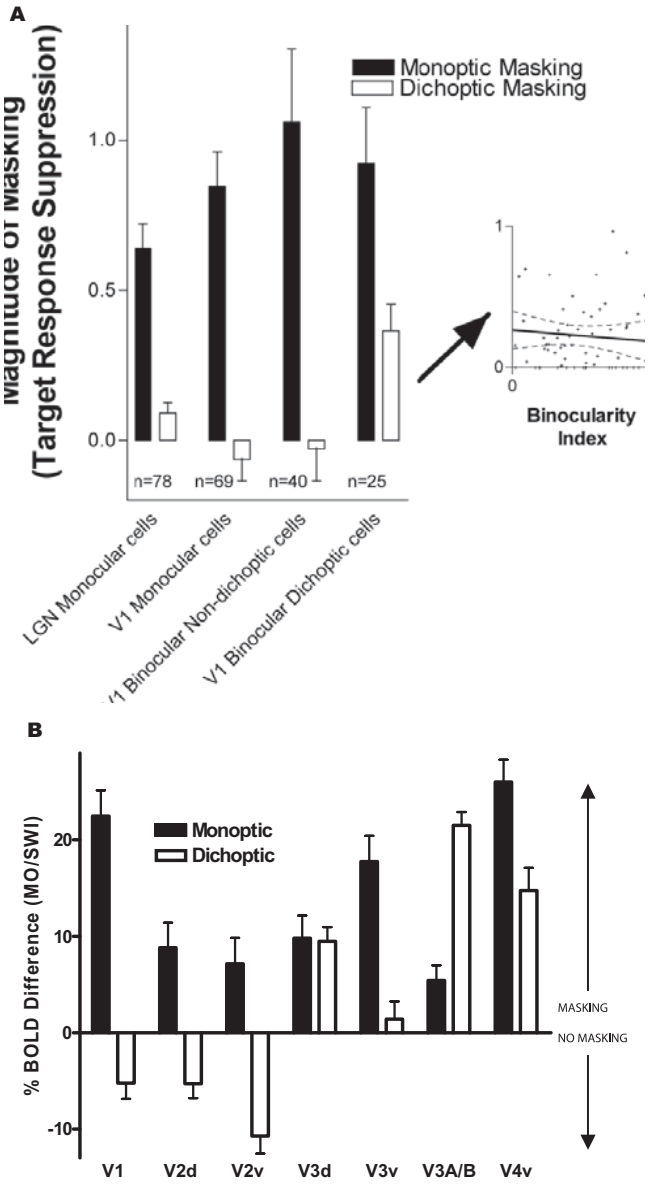


FIGURE 81.7 (*A*) Summary statistics of monoptic versus dichoptic masking responses in the LGN and area V1 of the macaque monkey. Monoptic (black bars) and dichoptic (white bars) masking magnitude as a function of cell type: LGN, V1 monocular, V1 binocular (nonresponsive to dichoptic masking), and V1 binocular (responsive to dichoptic masking) neurons. Inset shows the linear regression of dichoptic masking magnitude in V1 binocular neurons as a function of their degree of binocularity (all neurons plotted were significantly binocular as measured by their relative responses to monocular targets presented to the two eyes sequentially): BI of 0 indicates that the cells were monocular, while a BI of 1 means both eyes were equally dominant. (Reprinted from Macknik & Martinez-Conde, 2004a.) (*B*) Monoptic and dichoptic masking magnitude as a function of occipital retinotopic brain area in the human. Negative values indicate decreased visual masking (increased target visibility), whereas values ≥ 0 indicate increased masking (decreased target visibility). (Reprinted from Tse, Martinez-Conde, Schlegel, & Macknik, 2005.)

et al., 2005) (figure 81.7). There is no control condition with which to establish whether binocular rivalry fails because of a (mundane) lack of interocular suppression or (more interestingly) fails because of a lack of awareness maintaining circuits. One could address this issue by using binocular rivalry in tandem with a different stimulus to test for the explicit representation and strength of interocular suppression, such as visual masking stimuli. In visual masking, a monoptic form of the illusion is available, and so one can distinguish failures of interocular suppression from failures of visual awareness. But if one were to use masking stimuli to assess interocular suppression in a given visual area, then the role of such area in maintaining visibility and awareness would have also been established (in the absence of binocular rivalry stimulation), thus obviating the need for subsequent testing with binocular rivalry stimuli. Because one must rely on non-binocularly-rivalrous stimuli to determine the explicit representation and strength of interocular suppression in a given area, it is not possible to unambiguously interpret the neural correlates of perceptual state using binocular rivalry *alone* in any visual area, irrespective of the strength of the binocular rivalry response.

Our visual masking studies have shown that binocular neurons in areas V1 (the first stage in the visual hierarchy where information from the two eyes is combined) and V2 of humans and monkeys can integrate excitatory responses from the two eyes (Macknik & Martinez-Conde, 2004a; Tse et al., 2005) (figure 81.7). However, these same neurons do not express interocular suppression between the eyes. That is, binocular neurons in V1 are largely binocular for excitation while nevertheless being monocular for suppression (that is, input from one eye will not suppress the firing rate of a V1 or V2 cell that is primarily tuned for input from the opposite eye). Because most early binocular cells do not explicitly process interocular suppression, these neurons cannot explicitly process binocular rivalry. Further, binocular rivalry cannot distinguish between the role of interocular suppression and the role of awareness at any level of the visual system. Therefore no conclusions can be reached about the localization of the NCC to specific parts of the visual system based on binocular rivalry studies alone. If a given visual area does not correlate to binocular rivalry, it may simply mean that interocular suppression is not at play in that area, rather than that area is not maintaining awareness. However, these findings also beg the question of why some studies have found binocular rivalry modulation in low-level visual areas (Haynes, Driver, & Rees, 2005; Lee et al., 2005; Polonsky et al., 2000; Tong & Engel, 2001; Wunderlich, Schneider, & Kastner, 2005). One possible reason for this paradox is that these studies failed to control for the effects of attentional feedback, thus confounding apparent modulation to interocular suppression with attentional modulation. Because the subjects in these studies needed to attend to the binocular rivalry stimuli, attention itself, rather than binocular rivalry, may have produced the retinotopic activation found.

Monoptically and dichoptically presented visual-masking illusions (such as forward and backward masking and the standing wave) can differentiate between interocular suppression and awareness, and thus they are immune to these shortcomings. Therefore visual masking is an ideal illusion to isolate the NCC. Further, visual-masking illusions allow us to examine the brain's response to the same physical target under varying levels of visibility (unlike in binocular rivalry, where one only measures which of two rivalrous percepts was dominant at any given time, without consideration of how visible that percept was). Thus by quantifying the perceptual and physiological effects of visible versus invisible (masked) targets we will determine many, if not all, of the conditions that cause visibility.

We propose that, to test for explicit processing in neuronal populations or circuits, one should use a visual illusion, such as visual masking, that can be presented in at least two modes of operation: one mode to ensure that the neural circuit in question is able to process the stimulus explicitly and another mode to test the correlation to awareness. In visual masking, the monoptic mode establishes that the neural circuit in consideration explicitly processes visual-masking stimuli, and then the dichoptic mode can be used to probe the NCC.

Choosing an appropriate stimulus that is processed at multiple levels of the visual system is key to localizing awareness. However, one must take care to control for other potential experimental confounds. Lamme and colleagues used visual-masking stimuli to examine the NCC and concluded that stimulus-derived late responses (i.e., afterdischarges) are due to feedback from higher areas (Lamme, Zipser, & Spekreijse, 2002) and that this feedback is critical to maintaining awareness (figure 81.8). But if the late responses are due to feedback and not to feedforward circuits, their timing should be stable with respect to stimulus duration. That is, if late responses are due to feedback, target duration should not affect their latency, because the feedback would be driven by the target's onset-response as it rises through the visual hierarchy (figure 81.9*A*). On the contrary, if the late responses are caused by the target's termination in a feedforward fashion, then target duration would critically affect their latency (figure 81.9*B*). Figure 81.10 shows that, as the stimulus duration increases, so does the latency of the afterdischarge, against the predictions from Lamme's feedback model (Macknik & Martinez-Conde, 2004b). The model is further ruled out on psychophysical grounds, as the perceptual strength of masking varies with target duration (Macknik & Livingstone, 1998).

Despite these arguments, Lamme's group has maintained that late responses are due to feedback. In a recent study

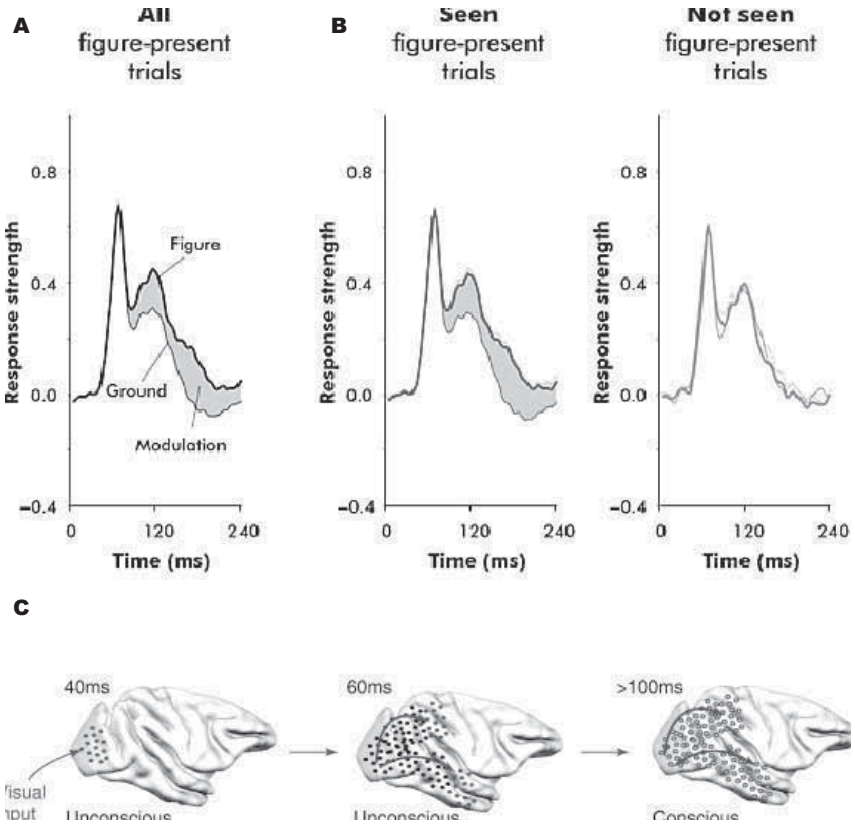


FIGURE 81.8 Alternative model of visual backward masking and awareness that require recurrent feedback. (*A*) Population response strength in awake monkey V1 to a figure-present (thick line) and ground (no figure-present, thin line) stimulus. (*B*) Responses to figure and ground conditions in which the figure was seen (left) and not seen (right). Notice that in the not-seen trials the late response does not differ for figure and ground conditions. (Reprinted from Super, Spekreijse, & Lamme, 2001.) (*C*) Model suggesting that the visibility-correlated late response in panels *A* and *B* is due to recurrent feedback from higher level cortices (activated by the feedforward onset response). (Reprinted from Lamme, 2003.)

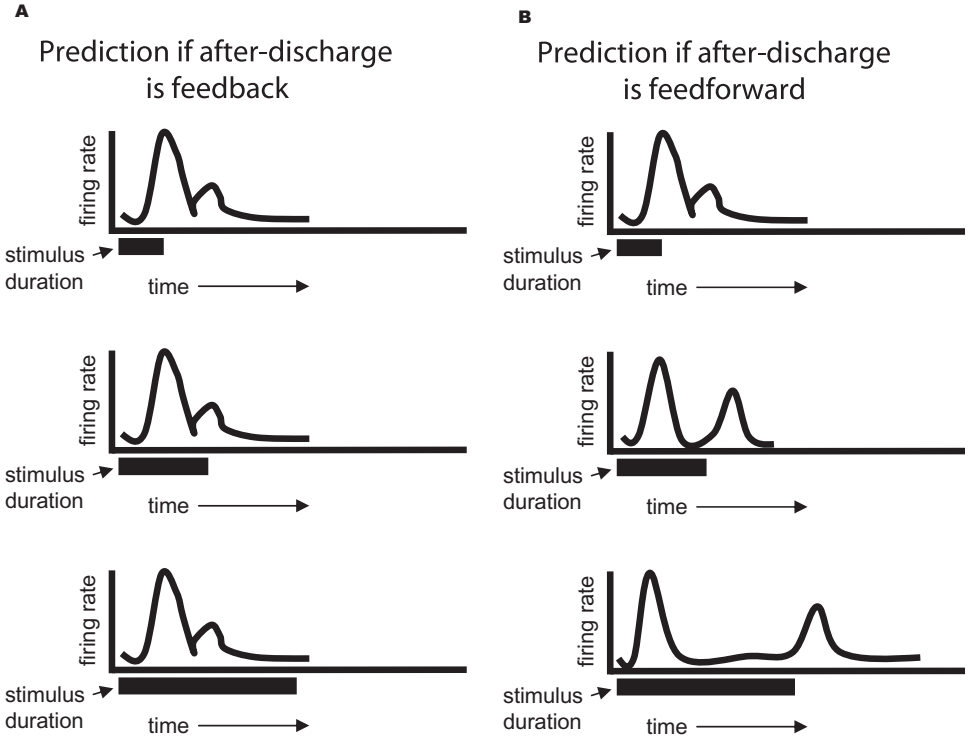


FIGURE 81.9 Predicted temporal dynamics of neuronal responses (as a function of stimulus duration) if the afterdischarge is (*A*) due to recurrent feedback driven by the stimulus onset or (*B*) driven by the termination of the stimulus in a feedforward manner.

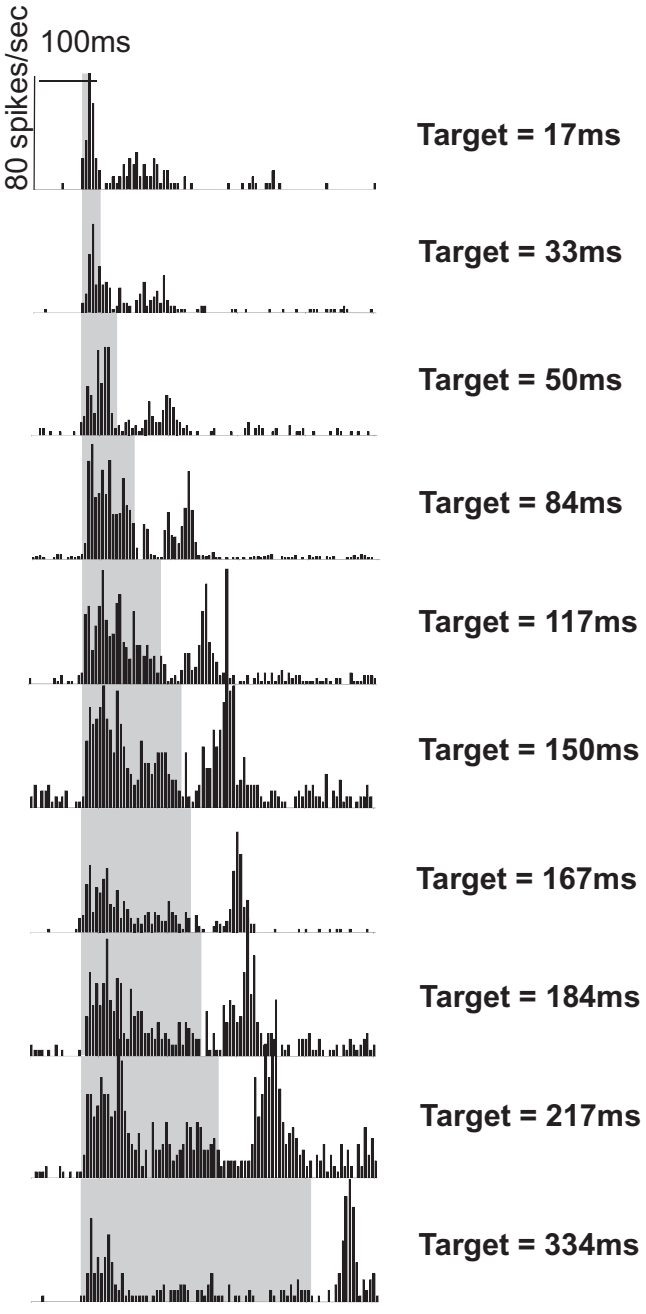


FIGURE 81.10 Typical responses from a single neuron in monkey area V1 to a target of various durations. The latency and magnitude of the afterdischarge grow as the target duration increases. (Reprinted from Macknik & Martinez-Conde, 2004a.)

they surgically removed the entire extrastriate visual cortex of a monkey (V3, V3A, V4, MT, MST, DP, LOP, LIPd, and 7a), a procedure which led to a reduction of area V1 late responses (Super & Lamme, 2007). However, surgical ablations are irreversible by definition: one cannot reverse the procedure to show that reinstating the ablated tissue cancels out the effect. Moreover, the surgical removal of the extrastriate cortex involves the resection of a large portion of the cerebral cortex, thus causing massive traumatic brain damage, including substantial damage to the cortical vascular systems as well as fibers of passage and nearby neural structures such as the optic radiations. Therefore it is unclear exactly what processes may or may not have been affected by such a drastic ablation.

Conclusions

We have reviewed the literature on the anatomy and physiology of feedback in the visual system and concluded that feedback connections may be the source of attentional facilitation and suppression, and that other proposed roles for feedback are not as clearly supported. We have also proposed that the large ratio of feedback to feedforward connections does not necessarily indicate a significant physiological role for feedback, but may instead be a requirement of any feedback pathway operating within a hierarchical neural system, such as the visual hierarchy. This statement would be true even if feedback subserves only a single role, such as top-down attentional modulation.

Finally, we have discussed the strengths of visual masking in the study of visual awareness, as compared to binocular rivalry, and have concluded that visual masking is a sound paradigm in awareness studies, whereas binocular rivalry has serious shortcomings as a tool to localize the NCC. Using visual masking as a tool, we have developed several new standards that must be met to determine the role of neural circuits, neurons, and brain areas in maintaining consciousness.

We have emphasized the need to control for the effects of attention as an important strategy in designing experiments that localize awareness. Attention can enhance or suppress the magnitude of neural responses to a given stimulus (Chen et al., 2008; Desimone & Duncan, 1995; McAdams & Maunsell, 1999; Moran & Desimone, 1985; Spitzer, Desimone, & Moran, 1988; Williford & Maunsell, 2006), and thus it may facilitate or suppress its perceptual awareness. However, attention is a distinct process from awareness itself (Koch & Tsuchiya, 2007; Merikle, 1980; Merikle & Joordens, 1997; Merikle, Smilek, & Eastwood, 2001). For instance, low-level bottom-up highly salient stimuli (such as flickering lights or loud noises) can lead to awareness and draw attention, even when the subject is actively attending to some other task, or not attending to anything (i.e., when the subject is asleep). It follows that experiments to isolate the NCC should control for the effects of attention.

Therefore, we add the following three standards for testing a neural circuit's contribution to awareness to Parker and Newsome's list:

8. The candidate neurons should be tested with an illusion that allows one to dissociate the physical stimulus from its perception. If the candidate set of neurons is capable of

maintaining awareness, the neural responses should match the subjective percept, rather than the objective physical stimulus.

9. The candidate neurons must explicitly process the type of information or stimulus used to test them.

10. The responses of the neurons, and of the perceiving subject, should be measured with experimental controls for the effect of attention.

ACKNOWLEDGMENTS We thank Mona Stewart, Hector Rieiro, and Jorge Otero-Millan for their technical assistance. This study was funded by the Barrow Neurological Foundation, National Science Foundation, Arizona Biomedical Research Institute, and Science Foundation Arizona.

REFERENCES

ALLMAN, J., MIEZIN, F., & MCGUINNESS, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu. Rev. Neurosci., 8*, 407–430.

ANGELUCCI, A., & BRESSLOFF, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog. Brain Res., 154*, 93–120.

ANGELUCCI, A., LEVITT, J. B., & LUND, J. S. (2002). Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area V1. *Prog. Brain Res., 136*, 373–388.

BARONE, P., BATARDIERE, A., KNOBLAUCH, K., & KENNEDY, H. (2000). Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J. Neurosci., 20*(9), 3263–3281.

BLASDEL, G. G., & LUND, J. S. (1983). Termination of afferent axons in macaque striate cortex. *J. Neurosci., 3*, 1389–1413.

BOUSSAOUD, D., UNGERLEIDER, L. G., & DESIMONE, R. (1990). Pathways for motion analysis: Cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *J. Comp. Neurol., 296*, 462–495.

BOYAPATI, J., & HENRY, G. (1984). Corticofugal axons in the lateral geniculate nucleus of the cat. *Exp. Brain Res., 53*, 335–340.

BULLIER, J., HUPE, J. M., JAMES, A., & GIRARD, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *J. Physiol. Paris, 90*(3–4), 217–220.

BULLIER, J., MCCOURT, M. E., & HENRY, G. H. (1988). Physiological studies on the feedback connection to the striate cortex from cortical areas 18 and 19 of the cat. *Exp. Brain Res., 70*, 90–98.

CHALMERS, D. J. (Ed.). (2000). *What is a neural correlate of consciousness?* Cambridge, MA: MIT Press.

CHEN, Y., MARTINEZ-CONDE, S., MACKNIK, S. L., BERESHPOLOVA, Y., SWADLOW, H. A., & ALONSO, J. M. (2008). Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nat. Neurosci., 11*(8), 974–982.

CRICK, F., & KOCH, C. (1995). Why neuroscience may be able to explain consciousness. *Sci. Am., 273*(6), 84–85.

CRICK, F., & KOCH, C. (1998a). Consciousness and neuroscience. *Cereb. Cortex, 8*(2), 97–107.

CRICK, F., & KOCH, C. (1998b). Constraints on cortical and thalamic projections—The no-strong-loops hypothesis. *Nature, 391*(6664), 245–250.

DESIMONE, R., & DUNCAN, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci., 18*, 193–222.

DESIMONE, R., SCHEIN, S. J., MORAN, J., & UNGERLEIDER, L. G. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Res., 25*, 441–452.

DOTY, R. W. (1983). Nongeniculate afferents to striate cortex in macaques. *J. Comp. Neurol., 218*(2), 159–173.

ERISIR, A., VAN HORN, S. C., & SHERMAN, S. M. (1997). Relative numbers of cortical and brainstem inputs to the lateral geniculate nucleus. *Proc. Natl. Acad. Sci. USA, 94*(4), 1517–1520.

FELLEMAN, D. J., & VAN ESSEN, D. C. (1991). Distributed hierarchal processing in the primate cerebral cortex. *Cereb. Cortex, 1*(1), 1–47.

FITZPATRICK, D., USREY, W. M., SCHOFIELD, B. R., & EINSTEIN, G. (1994). The sublaminar organization of corticogeniculate neurons in layer 6 of macaque striate cortex. *Vis. Neurosci., 11*(2), 307–315.

FRIES, W. (1990). Pontine projection from striate and prestriate visual cortex in the macaque monkey: An anterograde study. *Vis. Neurosci., 4*, 205–216.

FRIES, W., & DISTEL, H. (1983). Large layer VI neurons of monkey striate cortex (Meynert cells) project to the superior colliculus. *Proc. R. Soc. Lond. B Biol. Sci., 219*(1214), 53–59.

GIRARD, P., HUPE, J. M., & BULLIER, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J. Neurophysiol., 85*(3), 1328–1331.

GUILLERY, R. W. (1969). A quantitative study of synaptic interconnections in the dorsal lateral geniculate nucleus of the cat. *Z. Zellforsch, 96*, 39–48.

GUTIERREZ, C., & CUSICK, C. G. (1997). Area V1 in macaque monkeys projects to multiple histochemically defined subdivisions of the inferior pulvinar complex. *Brain Res., 765*(2), 349–356.

HAYNES, J. D., DRIVER, J., & REES, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron, 46*(5), 811–821.

HE, S., CAVANAGH, P., & INTRILIGATOR, J. (1996). Attentional resolution and the locus of visual awareness [see comments]. *Nature, 383*(6598), 334–337.

HENDRY, S. H., & YOSHIOKA, T. (1994). A neurochemically distinct third channel in the macaque dorsal lateral geniculate nucleus. *Science, 264*(5158), 575–577.

HILGETAG, C. C., O'NEILL, M. A., & YOUNG, M. P. (1996a). Indeterminate organization of the visual system. *Science, 271*(5250), 776–777.

HILGETAG, C. C., O'NEILL, M. A., & YOUNG, M. P. (1996b). On hierarchies. *Science, 271*(5250), 777b.

HUPE, J. M., JAMES, A. C., PAYNE, B. R., LOMBER, S. G., GIRARD, P., & BULLIER, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature, 394*(6695), 784–787.

KASTNER, S., NOTHDURFT, H. C., & PIGAREV, I. N. (1999). Neuronal responses to orientation and motion contrast in cat striate cortex. *Vis. Neurosci., 16*(3), 587–600.

KNIERIM, J. J., & VAN ESSEN, D. C. (1991). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol., 67*, 961–980.

KOCH, C., & TSUCHIYA, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends Cogn. Sci., 11*(1), 16–22.

LACHICA, E. A., & CASAGRANDE, V. A. (1992). Direct W-like geniculate projections to the cytochrome oxidase (CO) blobs in

primate visual cortex: Axon morphology. *J. Comp. Neurol., 319*(1), 141–158.

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci., 7*(1), 12–18.

Lamme, V. A., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *J. Cogn. Neurosci., 14*(7), 1044–1053.

Lee, S. H., Blake, R., & Heeger, D. J. (2005). Traveling waves of activity in primary visual cortex during binocular rivalry. *Nat. Neurosci., 8*(1), 22–23.

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry [see comments]. *Nature, 379*(6565), 549–553.

Livingstone, M. S., & Hubel, D. H. (1987). Connections between layer 4B of area 17 and the thick cytochrome oxidase stripes of area 18 in the squirrel monkey. *J. Neurosci., 7*(11), 3371–3377.

Lumer, E. D., Friston, K. J., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science, 280*(5371), 1930–1934.

Lund, J. S., Lund, R. D., Hendrickson, A. E., Bunt, A. H., & Fuchs, A. F. (1975). The origin of efferent pathways from the primary visual cortex, area 17, of the macaque monkey as shown by retrograde transport of horseradish peroxidase. *J. Comp. Neurol., 164*, 287–303.

Macknik, S. L. (2006). Visual masking approaches to visual awareness. *Prog. Brain Res., 155*, 179–217.

Macknik, S. L., & Haglund, M. M. (1999). Optical images of visible and invisible percepts in the primary visual cortex of primates. *Proc. Natl. Acad. Science USA, 96*(26), 15208–15210.

Macknik, S. L., & Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nat. Neurosci., 1*(2), 144–149.

Macknik, S. L., & Martinez-Conde, S. (2004a). Dichoptic visual masking reveals that early binocular neurons exhibit weak interocular suppression: Implications for binocular vision and visual awareness. *J. Cogn. Neurosci., 16*(6), 1–11.

Macknik, S. L., & Martinez-Conde, S. (2004b). The spatial and temporal effects of lateral inhibitory networks and their relevance to the visibility of spatiotemporal edges. *Neurocomputing, 58–60C*, 775–782.

Macknik, S., & Martinez-Conde, S. (2007). The role of feedback in visual masking and visual processing. *Advances in Cognitive Psychology, 3*(1–2), 125–152.

Macknik, S. L., & Martinez-Conde, S. (2009). Consciousness: Neurophysiology and visual awareness in. In L. R. Squire (Ed.), *Encyclopedia of neuroscience* (Vol. 3, pp. 105–116). Oxford, UK: Elsevier.

Macknik, S. L., Martinez-Conde, S., & Haglund, M. M. (2000). The role of spatiotemporal edges in visibility and visual masking. *Proc. Natl. Acad. Sci. USA, 97*(13), 7556–7560.

Martinez-Conde, S., Cudeiro, J., Grieve, K. L., Rodriguez, R., Rivadulla, C., & Acuna, C. (1999). Effects of feedback projections from area 18 layers 2/3 to area 17 layers 2/3 in the cat visual cortex. *J Neurophysiol., 82*(5), 2667–2675.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2000). Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. *Nat. Neurosci., 3*(3), 251–258.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2002). The function of bursts of spikes during visual fixation in the awake primate lateral geniculate nucleus and primary visual cortex. *Proc. Natl. Acad. Sci. USA, 99*(21), 13920–13925.

Maunsell, J. H., & Van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J. Neurosci., 3*, 2563–2586.

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci., 19*(1), 431–441.

Merikle, P. M. (1980). Selective metacontrast. *Can. J. Psychol., 34*(2), 196–199.

Merikle, P. M., & Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Conscious. Cogn., 6*(2–3), 219–236.

Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition, 79*(1–2), 115–134.

Mignard, M., & Malpeli, J. G. (1991). Paths of information flow through visual cortex. *Science, 251*, 1249–1251.

Montero, V. M. (1991). A quantitative study of synaptic contacts on interneurons and relay cells of the cat lateral geniculate nucleus. *Exp. Brain Res., 86*(2), 257–270.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782–784.

Murphy, P. C., & Sillito, A. M. (1996). Functional morphology of the feedback pathway from area 17 of the cat visual cortex to the lateral geniculate nucleus. *J. Neurosci., 16*(3), 1180–1192.

Myerson, J., Miezin, F., & Allman, J. M. (1981). Binocular rivalry in macaque monkeys and humans: A comparative study in perception. *Behav. Anal. Lett., 1*, 149–159.

Ogren, M., & Hendrickson, A. (1976). Pathways between striate cortex and subcortical regions in *Macaca mulatta* and *Saimiri sciureus*: Evidence for a reciprocal pulvinar connection. *Exp. Neurol., 53*(3), 780–800.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annu. Rev. Neurosci., 21*, 227–277.

Perkel, D. J., Bullier, J., & Kennedy, H. (1986). Topography of the afferent connectivity of area 17 in the macaque monkey: A double-labelling study. *J. Comp. Neurol., 253*, 374–402.

Peters, A., Payne, B. R., & Budd, J. (1994). A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cereb. Cortex, 4*(3), 215–229.

Polonsky, A., Blake, R., Braun, J., & Heeger, D. J. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nat. Neurosci., 3*(11), 1153–1159.

Rezak, M., & Benevento, L. A. (1979). A comparison of the organization of the projections of the dorsal lateral geniculate nucleus, the inferior pulvinar and adjacent lateral pulvinar to primary visual cortex (area 17) in the macaque monkey. *Brain Res., 167*(1), 19–40.

Rockland, K. S. (1996). Two types of corticopulvinar terminations: Round (type 2) and elongate (type 1). *J. Comp. Neurol., 368*(1), 57–87.

Rockland, K. S., Saleem, K. S., & Tanaka, K. (1994). Divergent feedback connections from areas V4 and TEO in the macaque. *Vis. Neurosci., 11*, 579–600.

Salin, P. A., Girard, P., Kennedy, H., & Bullier, J. (1992). Visuotopic organization of corticocortical connections in the visual system of the cat. *J. Comp. Neurol., 320*, 415–434.

Salin, P. A., Kennedy, H., & Bullier, J. (1995). Spatial reiprocity of connections between areas 17 and 18 in the cat. *Can. J. Physiol. Pharmacol., 73*(9), 1339–1347.

Sandell, J. H., & Schiller, P. H. (1982). Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *J. Neurophysiol., 48*, 38–48.

Sherman, S. M., & Guillery, R. W. (1998). On the actions that one nerve cell can have on another: Distinguishing "drivers" from "modulators." *Proc. Natl. Acad. Sci. USA, 95*(12), 7121–7126.

Sherman, S. M., & Guillery, R. W. (2002). The role of the thalamus in the flow of information to the cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci., 357*(1428), 1695–1708.

Shipp, S., & Zeki, S. (1989). The organization of connections between areas V5 and V1 in macaque monkey visual cortex. *Eur. J. Neurosci., 1*(4), 309–332.

Sillito, A. M., & Jones, H. E. (1996). Context-dependent interactions and visual processing in V1. *J. Physiol. Paris, 90*(3–4), 205–209.

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science, 240*, 338–340.

Super, H., & Lamme, V. A. (2007). Altered figure-ground perception in monkeys with an extrastriate lesion. *Neuropsychologia, 45*(14), 3329–3334.

Super, H., Spekreijse, H., & Lamme, V. A. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nat. Neurosci., 4*(3), 304–310.

Suzuki, W., Saleem, K. S., & Tanaka, K. (2000). Divergent backward projections from the anterior part of the inferotemporal cortex (area TE) in the macaque. *J. Comp. Neurol., 422*(2), 206–228.

Tong, F., & Engel, S. A. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature, 411*(6834), 195–199.

Troncoso, X. G., Macknik, S. L., & Martinez-Conde, S. (2005). Novel visual illusions related to Vasarely's "nested squares" show that corner salience varies with corner angle. *Perception, 34*(4), 409–420.

Troncoso, X. G., Tse, P. U., Macknik, S. L., Caplovitz, G. P., Hsieh, P. J., Schlegel, A. A., et al. (2007). BOLD activation varies parametrically with corner angle throughout human retinotopic cortex. *Perception, 36*(6), 808–820.

Tse, P. U., Martinez-Conde, S., Schlegel, A. A., & Macknik, S. L. (2005). Visibility, visual awareness, and visual masking of simple unattended targets are confined to areas in the occipital cortex beyond human V1/V2. *Proc. Natl. Acad. Sci. USA, 102*(47), 17178–17183.

Ungerleider, L. G., & Desimone, R. (1986a). Cortical connections of visual area MT in the macaque. *J. Comp. Neurol., 248*, 190–222.

Ungerleider, L. G., & Desimone, R. (1986b). Projections to the superior temporal sulcus from the central and peripheral field representations of V1 and V2. *J. Comp. Neurol., 248*, 147–163.

Van Horn, S. C., Erisir, A., & Sherman, S. M. (2000). Relative distribution of synapses in the A-laminae of the lateral geniculate nucleus of the cat. *J. Comp. Neurol., 416*(4), 509–520.

Vidyasagar, T. R., & Urbas, J. V. (1982). Orientation sensitivity of cat LGN neurones with and without inputs from visual cortical areas 17 and 18. *Exp. Brain Res., 46*, 157–169.

Wheatstone, C. (1838). On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions, 128*, 371–394.

Williford, T., & Maunsell, J. H. (2006). Effects of spatial attention on contrast response functions in macaque area V4. *J. Neurophysiol., 96*(1), 40–54.

Wunderlich, K., Schneider, K. A., & Kastner, S. (2005). Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat. Neurosci., 8*(11), 1595–1602.

Zeki, S. M. (1978a). Functional specialisation in the visual cortex of the rhesus monkey. *Nature, 274*(5670), 423–428.

Zeki, S. M. (1978b). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *J. Physiol., 277*, 273–290.

82 Emotion and Consciousness

MICHAEL KOENIGS AND RALPH ADOLPHS

ABSTRACT Emotional feelings, such as happiness and sadness, are a fundamental feature of human conscious experience, yet the neuroanatomical pathways that give rise to the conscious experience of emotion remain unclear. In this chapter we review data and theories that address the neural basis of the conscious experience of emotion. In particular, we highlight the insula's role in mapping the physiological state of the body as a potential substrate for the conscious experience of emotion. In addition, we discuss research supporting the assertion that emotional information need not be conscious to have significant effects on brain activity and behavior. We conclude with speculative thoughts about the relationship among interoception, emotion, and self-awareness.

"Consciousness" refers to subjective experience. At any given waking moment we are conscious *of* something. In other words, there are contents of consciousness—what it is that we are experiencing. The contents of consciousness can include things we see, hear, smell, taste, and touch. It is easy to become confused about what exactly it is that human conscious experience refers to, or whether such reference is veridical (see box 82.1 for some clarifications), but in this chapter we do not intend to discuss any of these issues. Instead, we focus on our commonsense understanding of what it means to be "conscious of" something, a sense that is shared in common between humans and many other animals. For conscious experiences in this everyday sense, there are solid neurobiological accounts of how a sensory stimulus (such as light) is processed by the nervous system: a specialized sensory apparatus (such as the retina) transduces physical energy to nerve impulses, and through afferent relays (such as thalamic nuclei) the information is mapped in modality-specific areas of the cerebral cortex (such as primary visual cortex). The cortical processing of sensory information is thought to give rise to the conscious experience of the sensation. The chapters in this volume by Koch (chapter 79) and Rees (chapter 80) address this issue in detail and discuss which cortical regions appear to contribute directly to the contents of consciousness. The neuroscience approach here has been straightforward: in which neural structures can we record signals (e.g., single-unit activity, BOLD response in fMRI studies) that correlate better with reports of conscious experience than with physical properties of the sensory stimuli alone?

But there are kinds of conscious experiences that seem different from the preceding examples and for which the neural pathways are less well understood. For example, the subjective experiences of joy and sorrow are no less real than the subjective experience of the blueness of the sky, but there seem to be more unanswered questions regarding the neurobiological mechanism for the experience of emotion than for the experience of vision. In the visual system the retina transduces light waves to neural impulses that code visual information—but is there an analogous point of origin for neural impulses that code emotional information? In the visual system there are visual cortices that are necessary for the conscious experience of sight—but are there analogous "emotional cortices" that are necessary for the conscious experience of emotion?

In this chapter we will review data and theories that address the neural basis of the conscious experience of emotion. In addition, we will discuss research supporting the assertion that emotional information need not be conscious to have significant effects on brain activity and behavior, just as is the case for nonconscious sensory processing. We will conclude with some more speculative thoughts on the relationship between emotion and consciousness, suggesting not only that emotion provides a possible content for consciousness, but that such content in fact provides an essential ingredient for all conscious experience.

Conscious experience of emotion

Before turning to a discussion of neuroscientific research relating emotion and consciousness, we first clarify some terminology. Depending on whom one asks—philosophers, psychologists, neuroscientists, or neurologists, for example—one may hear substantially different conceptions of both "emotion" and "consciousness." With respect to consciousness, we will make a distinction between the level (state) of consciousness and the contents of consciousness. The level of consciousness can be categorical (e.g., conscious versus unconscious) or graded (e.g., degrees of wakefulness and alertness); this is the aspect of consciousness that is impaired, as explained by Schiff in chapter 78 in this volume. As described previously, the contents of consciousness refer specifically to "what" it is that we are aware of in our conscious

MICHAEL KOENIGS University of Wisconsin, Madison, Wisconsin
RALPH ADOLPHS California Institute of Technology, Pasadena, California

Box 82.1 What are we conscious of?

A primary point of puzzlement, and material for a large literature in philosophy of mind and epistemology, is one illustrated in the following example. Suppose you are at this moment seeing a red sunset. What is it that you are directly conscious of? One answer, and the one we give as premise to this chapter, is to say that you are conscious of the sunset out there in the world. A different answer is to say that you are conscious of something in your mind or your brain. After all, you can experience a red sunset while dreaming or imagining it, and for all you know that could be the case even when you are convinced that you are really seeing it.

It is perhaps telling that no animal other than humans is capable of entertaining this doubt (we believe). It is also telling that nothing in the brain has the physical properties of being red or sunsetlike (although its representational properties may be such). While we can of course be mistaken about the state of the world on the basis of our conscious experience, we reject the extreme idea that we could be mistaken in a global and systematic way, a view called *solipsism*. In order for our conscious experiences to have any meaning, any content, they need to be anchored to the world in the general case (although, again, we can of course be mistaken on individual occasions—but such mistakes could only occur against a background of generally getting it right).

So what exactly is it that specific regions of the brain contribute to conscious experience? In our view, it is misleading to speak of regions of the brain as "having," "producing," or "causing" conscious experience. They contribute to it, certainly, but the way in which they do so should keep track of the right level at which to attribute conscious experience. When you experience the red sunset, no part of your brain has that experience—you, the person, does. The parts of your brain that contribute to the conscious experience all work together to make you conscious of the red sunset and able to tell us about it. We are, therefore, not conscious *of* neural activity, but rather conscious *with* it.

experience (a red sunset, a Bach concerto, a particular perfume, or an emotion). It is this aspect of consciousness (the contents of consciousness) whose relationship to emotion we will primarily discuss in this chapter. In this context we will distinguish between conscious and nonconscious processing of emotion, by which we mean whether or not an otherwise conscious individual (statewise) is specifically aware of particular emotional information (contentwise).

With respect to emotion, we will use a distinction between "emotion" and "feeling" that was articulated more than a century ago in the James-Lange theory of emotion (James, 1884; Lange, 1887) and later elaborated by Antonio Damasio (Damasio, 1994, 1999). In this framework, "emotion" refers to the physiological and musculoskeletal changes in the body (encompassing the body proper as well as physiological modulatory changes in the brain) that occur in response to an emotionally salient stimulus (e.g., changes in heart rate, posture, facial expression, alertness, and attention). "Feelings" refer to the conscious experience of emotion (e.g., the subjective experience of fear, sadness, or happiness). It follows that "emotions" can be observed objectively in the third person—for example, changes in heart rate can be measured with an EKG—whereas "feelings" are only available subjectively in the first person and thus can only be measured strictly through self-report (although they can of course be correlated with emotion and with neural responses). This usage of terms is by no means universal; while "feeling" is generally understood to mean the conscious experience of emotion, the usage of "emotion" is more vague and in many frameworks includes feelings. Furthermore, the word "feeling" has been used by many philosophers and scientists to denote any conscious, phenomenal state without necessarily implying emotional processing (e.g., "the feeling of seeing red"). In this chapter, we will use "feeling" to specifically denote the conscious experience of emotion, rather than the experience of conscious states in general. Figure 82.1 provides some further clarification of ways in which emotion, feeling, and related terms can be used.

A perennial debate about emotion and feeling as we have defined them is whether one causes the other. Intuitively, one might think that feelings cause emotions—you feel sad and consequently exhibit various autonomic and behavioral changes (you cry, etc.). William James famously inverted this intuitive causal order. According to James, we have autonomic emotional responses first, and these cause us to feel the emotion: you first run away from a bear and then feel afraid. Modern analyses of the relationship have acknowledged both versions. We now know that there can be very rapid emotional responses that precede feelings (and in fact can occur in the absence of our conscious awareness of the stimuli that triggered the emotion; see next subsection) in the way that James envisioned. But we also know that

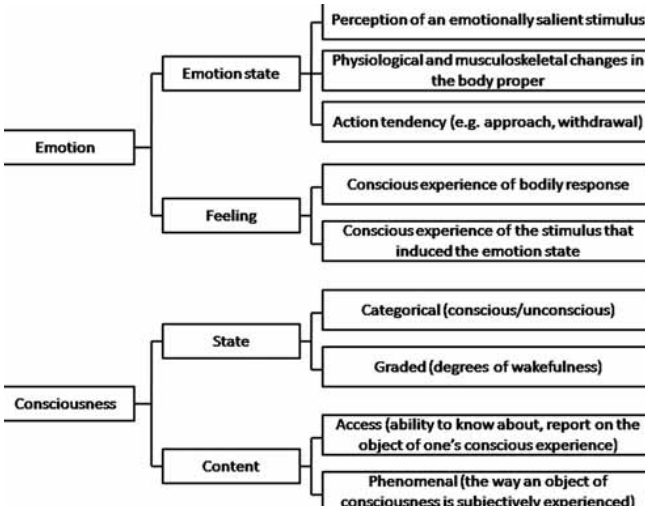


FIGURE 82.1 Clarification of terminology.

feelings influence emotions. The causal relationship is bidirectional.

Still, we ask the same question about feelings as about conscious experiences of other kinds: what are the necessary and sufficient neural substrates wherein neuronal activity covaries with the content of emotional conscious experience? We next review evidence for the idea that particular interoceptive sensory cortices are involved in feelings. Rather than making us aware of objects out there in the world (teloreception) or events at the interface of our body with the external environment (exteroception), interoception makes us aware of what is going on within the body proper (cf. box 82.1).

Somatic Changes as a Substrate for Feelings We think of physiological changes in the viscera and musculoskeletal system (which, together with changes in the brain itself, constitute emotions in our terminology) as the stimuli that determine the content of the conscious experience of emotion (in much the same way that light determines the content of the conscious experience of vision). Continuing the analogy with teloreceptive conscious experience, there must exist in the nervous system some mechanism for transducing these internal physiological changes into afferent neural impulses that are ultimately mapped in the brain in such a way that they generate the conscious experience of the emotion (the feeling). This type of afferent pathway does in fact exist; it has been dubbed the *interoceptive system* or the *homeostatic afferent system* (Craig, 2003) (figure 82.2). The essential function of this system is to monitor the physiological condition of the body. The peripheral nerve fibers of this system innervate virtually all tissues of the body, including skin, muscle, and internal organs. The specialized nerve endings of these fibers are sensitive to a variety of physiological parameters, such as temperature, mechanical stress, hormonal activity, pH, osmolarity, and metabolic activity. Neural impulses representing this physiological information ascend through a multisynaptic pathway. The peripheral fibers (small-diameter afferents) initially synapse in lamina I of the spinal and trigeminal dorsal horns. These cells project to the nucleus of the solitary tract and the parabrachial nucleus in the brain stem, as well as to the thalamus (specifically, the posterior part of the ventromedial nucleus of the thalamus, or VMpo, and the medial dorsal nucleus, or MD). The nucleus of the solitary tract and parabrachial nucleus make their own projections to the thalamus (specifically, the basal part of the ventromedial nucleus of the thalamus, or VMb, and MD). Thus three regions of the thalamus (VMpo, VMb, and MD) relay afferent homeostatic information to cortex. The primary cortical target of both VMpo and VMb is the dorsal insula. After the initial topographic projection from VMpo and VMb to dorsal insula, the afferent homeostatic information is remapped in the anterior insula. The primary cortical target of MD (the anterior cingulate cortex) also

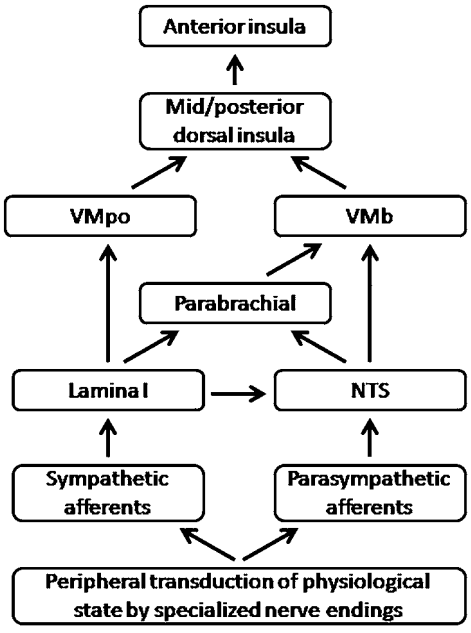


Figure 82.2 Schematic of homeostatic afferent system. NTS, nucleus of the solitary tract; VMpo, ventromedial thalamic nucleus (posterior portion); VMb, ventromedial thalamic nucleus (basal portion). (Adapted from Craig, 2003.)

projects to anterior insula. It is theorized that this cortical mapping (and remapping) of physiological/homeostatic information in anterior insula yields the conscious perceptions ("feelings") of bodily state, such as pain, hunger, and thirst (Craig, 2003). An additional and more speculative idea (Craig, 2003) is that awareness of one's own bodily feelings through this remapping may be a pathway unique to primates, or even *Homo sapiens*.

Inspired by the ideas of William James, Baruch Spinoza, and others (who were only able to conjecture at the time, ignorant of modern neuroscience findings), Antonio Damasio developed a theory of emotion and consciousness that highlights the recursive central mapping of peripheral physiological changes as the basis for subjective feeling and sentience (Damasio, 1994, 1999). Damasio's ideas predate those of Craig and in fact much neuroscience evidence; they were formulated in large part on his work with patients who had damage to the ventromedial prefrontal cortex—a region of the brain intimately connected with the insula and involved in interactions between emotion, decision making, and social behavior. Together, Damasio's and Craig's proposals relieve the question of a neural substrate for feelings from historical vagueness and provide specific hypotheses that it should depend on interoceptive representations.

Recent neuroscientific data support a basic tenet of Damasio's and Craig's theories, specifically, that a central dynamic mapping of the viscera in the anterior insula is critically involved in emotion and feeling. A variety of functional imaging studies associate activation of the anterior

insula with the subjective experience of emotion and/or homeostatic stress. Examples of stimuli inducing anterior insula activation include unpleasant taste (Small et al., 2003; Zald, Lee, Fluegel, & Pardo, 1998), disgusting smells (Heining et al., 2003; Wicker et al., 2003), pictures of disgusting food (Calder et al., 2007), capsaicin pain (Iadarola et al., 1998), thermal pain (Brooks, Zambreanu, Godinez, Craig, & Tracey, 2005), thirst (Farrell et al., 2006), air hunger (Evans et al., 2002), and unfair treatment (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Perhaps the most direct evidence linking the subjective feeling of emotion with the neurobiological mechanisms of interoception was provided by Hugo Critchley and colleagues (Critchley, Wiens, Rotshtein, Ohman, & Dolan, 2004). In this study, participants underwent functional magnetic resonance imaging while they attempted to determine whether their heartbeat was in sync with a series of auditory tones (a test of interoceptive awareness). Accuracy on the interoceptive heartbeat detection test was correlated with the activity and gray matter volume of the right anterior insula. Furthermore, participants' self-report ratings of day-to-day anxiety correlated with heartbeat detection accuracy and anterior insula activity. These findings suggest that anterior insula mediates both interoceptive awareness and emotional feeling states.

Just as higher-order visual cortices can mediate aspects of visual consciousness in the absence of a visual stimulus (e.g., in hallucinations, dreaming, and imagery), so can the insula mediate feelings that are decoupled from interoception. Its activity, and that of higher-order regions in the prefrontal cortex to which the insula projects, such as the anterior cingulate cortex, can be modulated by hypnosis (Rainville, Hofbauer, Bushnell, Duncan, & Price, 2002) and placebo effect (Petrovic, Kalso, Petersson, & Ingvar, 2002). If the insula is indeed critical for conscious emotional feelings, it is germane to consider how the properties of insula neurons may support this aspect of consciousness. In other domains of conscious experience (e.g., vision), researchers have theorized about the importance of thalamocortical loops, reciprocal feedback, and intercortical synchrony. Do insula neurons exhibit analogous profiles with respect to their connections with thalamus (e.g., VMpo and VMb) and other cortical areas (e.g., anterior cingulate)?

Although the mapping of emotion-related visceral changes in the insula is intuitively appealing and solidly grounded in neuroanatomical and physiological evidence, one must be wary of oversimplifying the neurobiological mechanism of feeling emotion. There is ample evidence to suggest that the mapping of visceral changes in the insula is neither necessary nor sufficient for the feeling of emotion. First of all, the basic James-Lange notion that emotional feelings necessarily follow from stereotyped changes in peripheral physiology (e.g., you feel sad because you cry) is challenged by reports of patients exhibiting pathological laughter and crying (PLC), which is a disorder of emotional expression in which the patient may laugh or cry (or rapidly switch between the two) in the absence of an appropriately motivating stimulus and without the affiliated feeling of happiness or sadness, respectively. Typically, PLC follows some neurological insult, such as tumor or cerebrovascular lesion. However, the lesions associated with PLC typically involve descending pathways in the internal capsule and brain stem, thus leaving intact the ascending pathways for representing peripheral physiological changes. The fact that salient and stereotyped changes in respiratory, orofacial, and muscular activity (laughing and crying) are entirely dissociable from the normally associated feelings indicates that one does not necessarily follow from the other. Second, it seems that the insula is involved in processing a limited range of emotion. As described earlier, the insula seems to respond preferentially to negatively valenced or noxious stimuli. In many studies insula activation is associated with basic motivational/appetitive states like hunger, thirst, and pain, which are typically considered more "physical" than primarily "affective" states. With respect to the basic emotions (e.g., fear, happiness, sadness, disgust, surprise, anger), insula activity appears to be most reliably associated with disgust, which has a prominent ingestive component that other basic emotions do not. Furthermore, there is no evidence indicating that focal lesions of the insula have pervasive effects on the experience of emotion. One study reports a reduction in the urge to smoke among patients with insula damage (Naqvi, Rudrauf, Damasio, & Bechara, 2007), but again, this is an effect on a motivational/appetitive state rather than a primarily emotional state. Thus the insula's role in the conscious experience of emotion may be restricted to negatively valenced or ingestion-related contexts. Human emotional experience, however, spans a much broader range. In the following sections, we discuss the neural architecture supporting more complex aspects of human emotional experience.

Emotion and Social Cognition One quintessential feature of human emotion is its prominence in social interaction. A subtle shift in gaze, posture, facial expression, or tone of voice by one person can engender marked changes in the emotional state of another. A fundamental question in the neuroscience of emotion, consciousness, and social behavior is how our conscious feeling state is impacted by the emotional states of others. An elementary step in this process is the recognition and discrimination of emotional states in others. Indeed, there has been a sustained effort to identify areas of the brain that discriminate between different expressions of emotion. For humans, facial expression is a principal means of communicating emotional state. Convergent lines of research implicate the amygdala as a key neural structure for the recognition of facial expressions

of emotion. The seminal report in this regard is that of SM, a woman who suffered extensive but remarkably selective damage to the amygdala bilaterally (Adolphs, Tranel, Damasio, & Damasio, 1994). SM evinces a marked impairment in the ability to recognize fear from facial expressions. Subsequent studies of patients with amygdala damage have replicated this deficit in the recognition of facial expressions of fear (Adolphs et al., 1999; Broks et al., 1998), but have also demonstrated deficits in the recognition of other basic emotions such as sadness, anger, and disgust (Adolphs & Tranel, 2004; Adolphs et al., 1999; Schmolck & Squire, 2001), as well as social emotions such as guilt, admiration, and flirtatiousness (Adolphs, Baron-Cohen, & Tranel, 2002) and the judgment of trustworthiness from faces (Adolphs, Tranel, & Damasio, 1998). Functional imaging studies of healthy individuals corroborate the conclusions of the patient studies; the amygdala is reliably activated during viewing of facial expressions of fear (Breiter et al., 1996; Morris et al., 1996). The amygdala's role in recognition of emotion from stimuli other than facial expressions is less clear. There are studies implicating the amygdala in recognition of emotion from auditory stimuli (Gosselin, Peretz, Johnsen, & Adolphs, 2007; Phillips et al., 1998) and body postures (de Gelder, Snyder, Greve, Gerard, & Hadjikhani, 2004; Hadjikhani & de Gelder, 2003), but there are also lesion studies that failed to find impairments in emotion recognition from such stimuli following amygdala lesions (Adolphs & Tranel, 1999; Atkinson, Heberlein, & Adolphs, 2007).

However, in order for one person's conscious feeling state to be impacted by the observed emotional state of another, the nervous system must do more than simply discriminate facial expressions of emotion. There must be an additional mechanism by which the observed emotion confers a change in the emotional state of the observer. One influential theoretical framework, known as *simulation theory*, holds that the observation and experience of certain emotions engage the same neural substrates. Simulation theory is based in part on electrophysiological recordings in monkeys, which have demonstrated the existence of brain cells ("mirror neurons") that fire both when the monkey performs a particular motor action and when a monkey observes the same action (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). As it relates to emotion, simulation theory has garnered empirical support through behavioral and functional imaging studies in humans. For example, Dimberg and colleagues demonstrated that when people are unconsciously exposed to emotional faces, they exhibit distinct facial muscle reactions that correspond to the subliminally presented emotional faces (Dimberg, Thunberg, & Elmehed, 2000). A complementary functional imaging study showed that some areas of the brain are activated both when a subject observes and consciously imitates facial expressions of emotion (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003). Furthermore, the anterior insula, which is known to be active during the experience of disgusting tastes and smells, is also active when viewing facial expressions of disgust in others (Phillips et al., 1997; Wicker et al., 2003). And Singer and colleagues found that certain areas of the brain, including anterior insula and rostral anterior cingulate cortex, are activated both when an individual experiences physical pain and when the individual is aware that a loved one is experiencing physical pain (Singer et al., 2004). Taken together, these studies support the notion that the neurobiological basis for human empathy relies, at least in part, on shared neural substrates for the experience and observation of emotion.

Nonconscious processing of emotion

In order to fully understand the neural underpinnings of the conscious experience of emotion, one must also consider the nonconscious aspects of emotional processing. There is ample evidence indicating that emotionally salient information in the environment can indeed have an impact on both

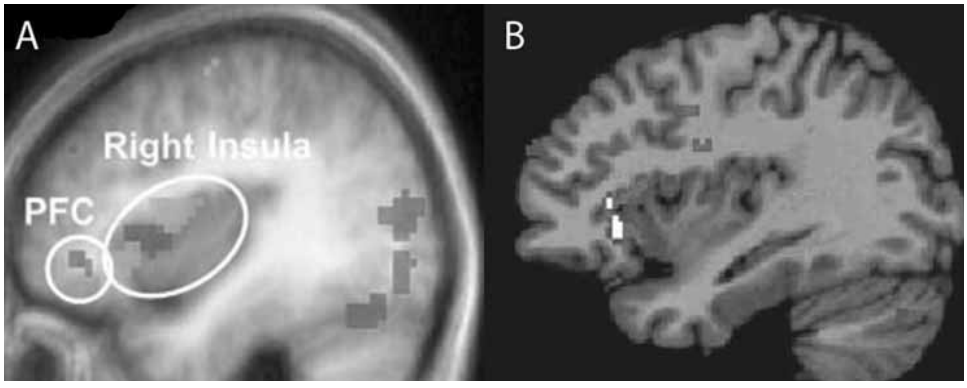


FIGURE 82.3 Functional MRI data supporting simulation theory. (*A*) Brain areas active during the observation of pain in another (red) and the feeling of pain in oneself (green) (Singer et al., 2004). (*B*) Brain areas active during the observation of disgust in another (blue) and the feeling of disgust in oneself (red); overlap in white (Wicker et al., 2003). (See color plate 101.)

brain and behavior without necessarily being consciously perceived.

Nonconscious Processing of Emotional Visual Stimuli Functional imaging studies have demonstrated measurable neural responses to emotional stimuli in the absence of conscious perception of the stimuli. As described previously, the amygdala is reliably activated when viewing facial expressions of emotion, particularly fear. Whereas initial studies displayed stimuli for at least several seconds to ensure conscious perception (Breiter et al., 1996; Morris et al., 1996), subsequent studies employed subliminal presentation paradigms to determine whether the amygdala's responsiveness to emotional faces was inextricably linked to conscious awareness of the faces. Using a backward masking procedure in which a briefly presented emotional face (fearful or happy) was immediately replaced by a neutral face, Whalen and colleagues found greater amygdala activation in response to the masked fearful faces compared to the masked happy faces, even when subjects reported being unaware of the emotional faces (Whalen et al., 1998). In studies that suppressed the conscious perception of presented faces (Jiang & He, 2006; Williams, Morris, McGlone, Abbott, & Mattingley, 2004), the amygdala was again activated for fearful faces even when the faces were not consciously perceived. To determine the neural pathway that mediates the amygdala's response during the subliminal presentation of fearful faces, Morris and colleagues examined functional connectivity data in an fMRI study of backward-masked fear-conditioned faces (Morris, Ohman, & Dolan, 1999). In response to unseen faces (compared to seen faces), there was increased connectivity between the amygdala and the subcortical visual system (pulvinar and superior colliculus) and decreased connectivity between the amygdala and cortical visual areas (e.g., fusiform gyrus). These results suggest that the amygdala's response to nonconcious emotionally salient information is mediated by a subcortical (colliculothalamic) pathway. This conclusion is supported by the study of a patient (GY) who is blind in half of his visual field as a result of cortical damage involving primary visual areas. Despite his report of not being able to consciously "see" any faces in his blind hemifield, GY is able to discriminate, above chance, the faces' emotional expressions (de Gelder, Vroomen, Pourtois, & Weiskrantz, 1999). Furthermore, GY exhibits amygdala activity in response to fearful and fear-conditioned faces presented in his blind hemifield, and this amygdala activity correlates with activity in his intact subcortical pathway.

Autonomic Responses as Indices of Nonconscious Emotion Psychophysiological and behavioral studies provide further evidence for nonconscious emotional processing. Skin conductance response (SCR) is an index of autonomic nervous system activity that has been widely used as a measure of emotional arousal. The advantage of this technique for measuring emotional arousal is that it is "objective" in the sense that it does not depend on a subjective self-report. This feature of SCR is particularly useful for assessing emotional responsiveness in certain types of brain-damaged patients, who may lose access to declarative knowledge (and conscious self-report) for certain classes of stimuli or in certain contexts. For example, damage to occipitotemporal cortex may result in a condition known as prosopagnosia, which is an inability to recognize previously familiar faces, such as friends or family, despite intact basic visual perceptual abilities. Tranel and Damasio studied a group of six prosopagnosics to determine whether the patients' SCRs could discriminate between familiar and unfamiliar faces even if the patients could not discriminate the faces at a conscious level (Tranel & Damasio, 1985). This hypothesized dissociation is exactly what Tranel and Damasio found; although none of the prosopagnosics had any conscious sense of familiarity for any of the faces, each patient produced significantly larger SCRs for the faces of well-known family and friends than for the faces of strangers.

In addition, SCR data have been used to demonstrate that emotional responses can be acquired independently of conscious awareness of the emotion-inducing stimuli. In one of a series of such studies, Ohman and Soares collected SCR data from spider- and snake-phobic individuals as they were subliminally presented backward-masked pictures of spiders and snakes (Ohman & Soares, 1994). Despite reporting no conscious awareness of the phobia-relevant pictures, the phobic subjects still produced large SCRs. In a related study, Ohman and Soares tested nonphobic individuals on a conditioning paradigm in which masked pictures of snakes and spiders were repeatedly paired with an electric shock (Ohman & Soares, 1998). Again, despite no conscious recognition of the conditioned stimuli, subjects developed large conditioned SCRs.

The independence of emotional responsiveness from conscious awareness has been further supported by lesion patient data. Bechara and colleagues studied a patient with dense anterograde amnesia following bilateral hippocampal damage, using a conditioning paradigm in which a previously "neutral" stimulus (a particular visual slide or auditory tone) was repeatedly paired with a sudden, loud burst of noise (Bechara et al., 1995). Like normal healthy adults, the amnesic patient developed large SCRs in response to the conditioned stimuli. But unlike healthy adults, who could readily describe the association between the conditioned and unconditioned stimuli, the amnesic patient had no conscious knowledge of the association.

More elaborate studies of brain-injured patients have highlighted the role of nonconscious emotional processes in

overt behavior. One such study (Tranel & Damasio, 1993) featured patient B, who suffered severe amnesia with prosopagnosia following extensive bilateral cortical and subcortical damage resulting from herpes simplex encephalitis. Anecdotal observation of patient B suggested that he had a preference for a particular caregiver, despite having no declarative memory of the person, such as name, appearance, personality characteristics, hours of work, or specific interactions. To formally test the idea that patient B had developed a nonconscious affective behavioral preference, Tranel and Damasio devised an experiment in which patient B was exposed to three different caregivers: a "good guy," who was friendly and always granted requests; a "bad guy," who was responsible for tedious and unpleasant tasks; and a "neutral guy," who was neither overly friendly nor unpleasant. After a week of exposure to the different caregivers, patient B denied any conscious recognition of any of the three persons and could offer no declarative information about any of them, even with extensive prompting. However, in a two-alternative forced-choice test in which a picture of each caregiver was paired with a picture of a visually similar stranger, patient B exhibited a strong preference for the "good guy," a strong aversion for the "bad guy," and no preference or aversion for the "neutral guy." In addition, patient B had much larger SCRs for the "good guy" and the "bad guy" than for the "neutral guy" and strangers. These data clearly demonstrate that affective behavioral preference is not necessarily mediated by conscious recognition of the alternatives.

Nonconscious Emotion in Decision Making The conclusion concerning affective preference unmediated by conscious awareness is further supported by a series of studies on the Iowa Gambling Task, or IGT (Bechara, Damasio, Damasio, & Anderson, 1994). In this task, the subject must choose between one of four decks of cards, with each card denoting a net gain or loss of money. The key manipulation is that the decks are preprogrammed so two decks are "good" in the long run, resulting in net gain, and two decks are "bad" in the long run, resulting in net loss. Through the variable experiences of reward and punishment with each deck, most normal people eventually develop a strong preference for the good decks. Bechara and colleagues later demonstrated that over the course of the task normal subjects develop SCRs prior to card selection, and that these "anticipatory SCRs" were greater when preceding choices for the bad decks (Bechara, Tranel, Damasio, & Damasio, 1996). In a subsequent study, Bechara and colleagues intermittently queried subjects throughout the task to assess the subjects' conscious declarative knowledge of the decks' relative net values (Bechara, Damasio, Tranel, & Damasio, 1997). Intriguingly, some normal subjects were never able to explicitly describe the relative long-term outcomes of each deck, but still generated anticipatory SCRs and chose advantageously. Conversely, several patients with blunted emotional responsiveness due to ventromedial prefrontal lesions never generated anticipatory SCRs and continued to choose disadvantageously throughout the task, even in cases where they could articulate the relative net values of each deck. Thus the critical determinant of affective preference may not be conscious consideration of likely outcomes, but a nonconscious somatic signal derived from prior experiences with reward and punishment.

The assertion of the importance of nonconscious (or at least not necessarily conscious) somatic signals in decision making has been debated. Subsequent studies of the IGT have directly addressed the issue of conscious and nonconscious processing. In a study that involved more sensitive and detailed assessments of conscious knowledge of deck contingencies during the task, Maia and McClelland found that subjects were able to provide task-relevant quantitative knowledge earlier in the task than Bechara and colleagues had claimed, and furthermore that this conscious knowledge correlated positively with task performance (Maia & McClelland, 2004). In addition, a patient study shows that amnesics with impaired declarative recall typically perform the IGT at chance levels (Gutbrod et al., 2006). These findings suggest that conscious processing of previous reward/punishment experiences may indeed be the critical determinant of IGT performance. However, a later study introduced a manipulation in which subjects were explicitly told that the good and bad decks were reversed at one point during the task, and after that point there would be no ongoing feedback of gains and losses (Stocco & Fum, 2008). If subjects had conscious knowledge of the deck contingencies at the point of reversal, they would switch their selections accordingly. Despite demonstrating a behavioral preference for the good decks prior to the explicit reversal, subjects did not reliably shift card selection following the reversal, thereby indicating that conscious explicit knowledge of relative deck values was not the crucial determinant of successful task performance.

A recent study proposes a reconciliation of these apparently conflicting results (Persaud, McLeod, & Cowey, 2007). In this study, awareness of relative deck values was assessed with a postselection wagering scheme. In other words, subjects could make a side bet on whether each selection would pay off. If subjects were consciously aware that a deck was likely to yield a positive result, they would wager more money on a selection from that deck. Persaud and colleagues observed that when using Bechara's testing scheme, which involved intermittent open-ended questions about the game, the subjects' wagering gains lagged behind their IGT gains, indicating that subjects were improving on the IGT without conscious awareness of deck contingencies. However, when Persaud and colleagues used Maia and McClelland's

testing scheme, which involved detailed quantitative questions about the game, the subjects' wagering gains matched their IGT gains, indicating that IGT performance was coupled with conscious knowledge about the game. These results suggest that conscious awareness is not necessary for improved decision making on the IGT, but that conscious awareness can be induced by directed questioning, and that this increased awareness is beneficial for performance.

Summary and closing thoughts

We finish by returning to ideas articulated by Damasio in his writings. According to this view, the body provides the substrate for conscious experience through its interaction with external objects. The self (in the sense of the material organism) is represented by "first-order" maps of afferent homeostatic activity, which are impacted by external objects and/or mental images. The physiological changes comprising emotion are one source of fluctuation in the first-order maps. In turn, higher-order maps in the brain represent the act of first-order somatic mapping by the nervous system. In this way, the organism becomes aware of itself interacting with the environment and experiencing emotion. The result is a conscious experience associated with a sense of self. The brain stem nuclei and white matter tracts that constitute the initial stages of all interoceptive channels are therefore also a prime region for which to hypothesize deficits in consciousness following lesions. While the impairments in consciousness (e.g., coma, persistent vegetative state) that can result from brain stem lesions are usually interpreted as arising from damage to ascending arousal systems, Damasio provides an alternative idea. The upper brain stem (at or above the level at which somatic information from the head is conveyed to the trigeminal nucleus) is the only place where a single lesion could interrupt all ascending interoceptive information (cervical spinal cord transection, for instance, still leaves intact information from the head coming in through the trigeminal nerve, as well as information from the body by way of the vagus nerve). In this view, the complete loss of ascending interoceptive information would render one unconscious.

REFERENCES

Adolphs, R., Baron-Cohen, S., & Tranel, D. (2002). Impaired recognition of social emotions following amygdala damage. *J. Cogn. Neurosci.*, *14*(8), 1264–1274.

Adolphs, R., & Tranel, D. (1999). Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia*, *37*(11), 1285–1292.

Adolphs, R., & Tranel, D. (2004). Impaired judgments of sadness but not happiness following bilateral amygdala damage. *J. Cogn. Neurosci.*, *16*(3), 453–462.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*(6684), 470–474.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669–672.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., et al. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*(10), 1111–1117.

Atkinson, A. P., Heberlein, A. S., & Adolphs, R. (2007). Spared ability to recognise fear from static and moving whole-body cues following bilateral amygdala damage. *Neuropsychologia*, *45*(12), 2772–2782.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1–3), 7–15.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science*, *269*(5227), 1115–1118.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cereb. Cortex*, *6*(2), 215–225.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, *17*(5), 875–887.

Broks, P., Young, A. W., Maratos, E. J., Coffey, P. J., Calder, A. J., Isaac, C. L., et al. (1998). Face processing impairments after encephalitis: Amygdala damage and recognition of fear. *Neuropsychologia*, *36*(1), 59–70.

Brooks, J. C., Zambreanu, L., Godinez, A., Craig, A. D., & Tracey, I. (2005). Somatotopic organisation of the human insula to painful heat studied with high resolution functional imaging. *NeuroImage*, *27*(1), 201–209.

Calder, A. J., Beaver, J. D., Davis, M. H., van Ditzhuijzen, J., Keane, J., & Lawrence, A. D. (2007). Disgust sensitivity predicts the insula and pallidal response to pictures of disgusting foods. *Eur. J. Neurosci.*, *25*(11), 3422–3428.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. USA*, *100*(9), 5497–5502.

Craig, A. D. (2003). Interoception: The sense of the physiological condition of the body. *Curr. Opin. Neurobiol.*, *13*(4), 500–505.

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nat. Neurosci.*, *7*(2), 189–195.

Damasio, A. R. (1994). *Descartes' error*. New York: Penguin.

Damasio, A. R. (1999). *The feeling of what happens*. New York: Harcourt.

de Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjikhani, N. (2004). Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proc. Natl. Acad. Sci. USA*, *101*(47), 16701–16706.

de Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. *NeuroReport*, *10*(18), 3759–3763.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.*, *11*(1), 86–89.

Evans, K. C., Banzett, R. B., Adams, L., McKay, L., Frackowiak, R. S., & Corfield, D. R. (2002). BOLD fMRI identifies limbic, paralimbic, and cerebellar activation during air hunger. *J. Neurophysiol.*, *88*(3), 1500–1511.

Farrell, M. J., Egan, G. F., Zamarripa, F., Shade, R., Blair-West, J., Fox, P., et al. (2006). Unique, common, and interacting cortical correlates of thirst and pain. *Proc. Natl. Acad. Sci. USA*, *103*(7), 2416–2421.

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, *45*(2), 236–244.

Gutbrod, K., Krouzel, C., Hofer, H., Muri, R., Perrig, W., & Ptak, R. (2006). Decision-making in amnesia: Do advantageous decisions require conscious knowledge of previous behavioural choices? *Neuropsychologia*, *44*(8), 1315–1324.

Hadjikhani, N., & de Gelder, B. (2003). Seeing fearful body expressions activates the fusiform cortex and amygdala. *Curr. Biol.*, *13*(24), 2201–2205.

Heining, M., Young, A. W., Ioannou, G., Andrew, C. M., Brammer, M. J., Gray, J. A., et al. (2003). Disgusting smells activate human anterior insula and ventral striatum. *Ann. NY Acad. Sci.*, *1000*, 380–384.

Iadarola, M. J., Berman, K. F., Zeffiro, T. A., Byas-Smith, M. G., Gracely, R. H., Max, M. B., et al. (1998). Neural activation during acute capsaicin-evoked pain and allodynia assessed with PET. *Brain*, *121*(Pt. 5), 931–947.

James, W. (1884). What is an emotion? *Mind*, *9*, 188–205.

Jiang, Y., & He, S. (2006). Cortical responses to invisible faces: Dissociating subsystems for facial-information processing. *Curr. Biol.*, *16*(20), 2023–2029.

Lange, C. (1887). *Ueber Gemuthsbewegungen*, *3*, 8.

Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa Gambling Task. *Proc. Natl. Acad. Sci. USA*, *101*(45), 16075–16080.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, *383*(6603), 812–815.

Morris, J. S., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc. Natl. Acad. Sci. USA*, *96*(4), 1680–1685.

Naqvi, N. H., Rudrauf, D., Damasio, H., & Bechara, A. (2007). Damage to the insula disrupts addiction to cigarette smoking. *Science*, *315*(5811), 531–534.

Ohman, A., & Soares, J. J. (1994). "Unconscious anxiety": Phobic responses to masked stimuli. *J. Abnorm. Psychol.*, *103*(2), 231–240.

Ohman, A., & Soares, J. J. (1998). Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following nonrecognized fear-relevant stimuli. *J. Exp. Psychol. Gen.*, *127*(1), 69–82.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nat. Neurosci.*, *10*(2), 257–261.

Petrovic, P., Kalso, E., Petersson, K. M., & Ingvar, M. (2002). Placebo and opioid analgesia—imaging a shared neuronal network. *Science*, *295*(5560), 1737–1740.

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proc. Biol. Sci.*, *265*(1408), 1809–1817.

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*(6650), 495–498.

Rainville, P., Hofbauer, R. K., Bushnell, M. C., Duncan, G. H., & Price, D. D. (2002). Hypnosis modulates activity in brain structures involved in the regulation of consciousness. *J. Cogn. Neurosci.*, *14*(6), 887–901.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.*, *3*(2), 131–141.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, *300*(5626), 1755–1758.

Schmolck, H., & Squire, L. R. (2001). Impaired perception of facial emotions following bilateral damage to the anterior temporal lobe. *Neuropsychology*, *15*(1), 30–38.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*(5661), 1157–1162.

Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M., & Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron*, *39*(4), 701–711.

Stocco, A., & Fum, D. (2008). Implicit emotional biases in decision making: The case of the Iowa Gambling Task. *Brain Cogn.*, *66*(3), 253–259.

Tranel, D., & Damasio, A. R. (1985). Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics. *Science*, *228*(4706), 1453–1454.

Tranel, D., & Damasio, A. R. (1993). The covert learning of affective valence does not require structures in the hippocampal system or amygdala. *J. Cogn. Neurosci.*, *5*, 79–88.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J. Neurosci.*, *18*(1), 411–418.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655–664.

Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *J. Neurosci.*, *24*(12), 2898–2904.

Zald, D. H., Lee, J. T., Fluegel, K. W., & Pardo, J. V. (1998). Aversive gustatory stimulation activates limbic circuits in humans. *Brain*, *121*(Pt. 6), 1143–1154.

83 Volition and the Function of Consciousness

HAKWAN LAU

ABSTRACT What are the psychological functions that can only be performed consciously? People have intuitively assumed that many acts of volition are not influenced by unconscious information. These acts range from simple examples such as making a spontaneous motor movement to higher cognitive control. However, the available evidence suggests that under suitable conditions, unconscious information can influence these behaviors and the underlying neural mechanisms. One possibility is that stimuli that are consciously perceived tend to yield strong signals in the brain, which makes us think that consciousness has the function of such strong signals. However, if we could create conditions where the stimuli could yield strong signals but not the conscious experience of perception, perhaps we would find that such stimuli are just as effective in influencing volitional behavior. Future studies that focus on clarifying this issue may tell us what the defining functions of consciousness are.

Many acts of volition seem to require conscious effort. We consciously initiate spontaneous motor movements. We cancel planned actions at will. We deliberately avoid particular actions. We intentionally shift our action plans in order to pursue different goals. Sometimes, theorists say, these are the functions of consciousness, as if evolution has equipped us with the gift of consciousness just to perform these acts. Without consciousness, presumably, we would only be able to perform much simpler actions that are no more sophisticated than embellished reflexes.

In this chapter we review available evidence to see if these intuitive claims are empirically supported. Recent studies in cognitive neuroscience suggest that many of these complex processes can actually be performed without consciousness. Or at least, many of them can be directly influenced by unconscious information. This evidence calls into question the true function of consciousness, if it is not to enable us to deliberate our actions. We end by discussing what is logically required for an experiment to demonstrate the true function of consciousness.

Spontaneous motor initiation

Motor actions that are made not in immediate or direct response to external stimuli can be said to be spontaneously initiated. These are also sometimes called self-paced or self-generated actions. For instance, one may choose to casually flex one's wrist while sitting in a dark room, out of one's own free choice and timing, not to react to anything in particular. Some philosophers have argued that in cases like that, it should seem obvious that the action is caused by one's conscious intention (Searle, 1983). Whereas one may argue that fast reactions to external stimuli may be driven by unconscious reflex (e.g., a runner leaping forward upon hearing the starting shot), spontaneous actions do not seem to have any immediate cause but the conscious intention itself.

However, it has been shown that there is preparatory activity in the brain that starts at as early as 1–2 seconds before spontaneous actions are executed. This piece of one of the most perplexing findings in cognitive neuroscience was originally reported by Kornhuber and Deecke in the 1960s (Kornhuber & Deecke, 1965). They placed electrodes on the scalp to measure electroencephalography (EEG) while subjects made spontaneous movements at their own timing. The EEG data that were time-locked to the point of motor execution (as measured by muscle contraction indicated by electromyography, EMG) were averaged over many trials, thereby producing an event-related potential (ERP) known as the *Bereitschaftspotential* (BP) or readiness potential (RP). The readiness potential is slowly rising, peaking at around the point of action execution, and starting from 1 to 2 seconds before that (figure 83.1). The readiness potential is most pronounced at electrodes near the vertex (Cz in the EEG coordinate system), which is directly above the medial premotor areas (including the supplementary motor area, SMA, presupplementary motor area, pre-SMA, and cingulate motor areas below them). It is generally believed that one major source of the readiness potential lies in the medial premotor areas (Ball et al., 1999; Erdler et al., 2000; Weilke et al., 2001; Cunnington, Windischberger, Deecke, & Moser, 2003).

The demonstration of the readiness potential calls into question whether spontaneous movements are really caused by the preceding conscious intentions. Intuitively, conscious intentions seem to cause motor actions almost immediately—it seems to take much less time than 1–2 seconds. This could mean that the brain starts to prepare for the actions long before we consciously initiate them.

HAKWAN LAU Columbia University, New York, New York

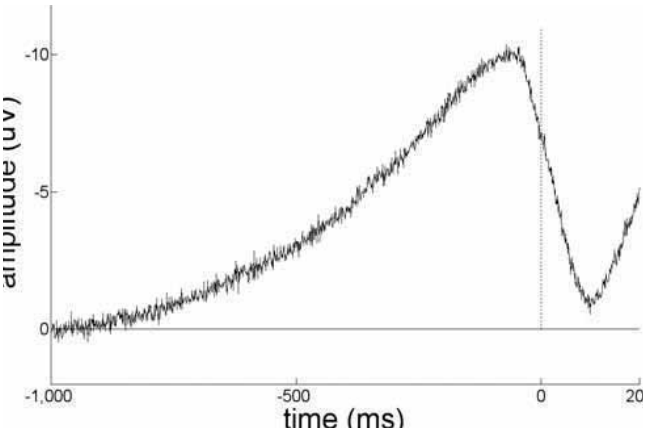


Figure 83.1 A schematic depiction of the readiness potential (RP) preceding spontaneous movements. The RP is usually recorded at the top of the scalp, above medial frontal premotor areas. It gradually ramps up, beginning about 1–2 seconds before movement and peaking around the time of movement execution (marked as time = 0).

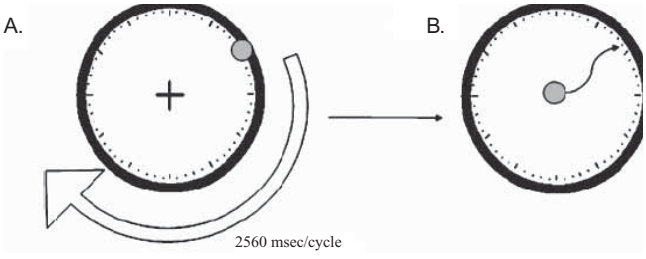


Figure 83.2 The Libet clock paradigm. (*A*) The subject views a dot rotating slowly (2.56 seconds per cycle) around a clock face and waits for an urge to move to arise spontaneously. When the urge arrives, the subject makes a movement (e.g., a key press). (*B*) After making the movement, the subject estimates the earliest time at which the intention to move was experienced. To carry out this time estimate, the subject either verbally indicates the location of the dot where the intention was first felt or moves a cursor to that location (as in this example). In a common control condition, the subject uses the clock to estimate the time of movement rather than the onset of intention. (Edited and adapted from Lau, Rogers, & Passingham, 2007.)

Benjamin Libet and colleagues empirically studied the timing of the conscious intention in relation to the readiness potential and the action (Libet, Gleason, Wright, & Pearl, 1983). To measure the onset of conscious intention, he invented a creative but controversial paradigm that is sometimes called the *Libet clock paradigm*. In those studies, subjects watched a dot revolving around a clock face at a speed of 2.56 seconds per cycle, while they flexed their wrist spontaneously (figure 83.2). After the action was finished, subjects were required to report the location of the dot when they "first felt the urge" to produce the action, that is, the onset of intention. The subject might say it was at the 3 o'clock or 4 o'clock position when they first felt the intention, for instance. This way the subjects could time and report the onset of their intention, and the experimenter could then work out when the action was actually produced, and hence the temporal distance between the two. Libet and colleagues reported that subjects on average report the onset of intention to be about 250 ms before motor execution.

Many people feel uncomfortable with the fact that the onset of the readiness potential seems to be so much earlier than the onset of intention, and some have tried to explain away the gap. Libet and colleagues have tried to study the onset of the readiness potential more carefully, discarding trials that might have been "contaminated" by preplanning of action well before the action (for instance, by counting to 10 and then triggering the movement), as reported by the subjects. By only looking at the trials where the actions were supposed to be genuinely spontaneous, Libet and colleagues reported that the onset of the readiness potential is only about 500 ms before action execution (Libet, Gleason, et al., 1983). However, this is still clearly earlier than the reported onset of intention. And by discarding so many trials, it may be that the analysis just lacked the power to detect an earlier onset.

Some have argued that the onset of readiness potential might be an artifact due to the averaging needed to produce the ERP (Miller & Trevena, 2002). However, Romo and Schultz (1987) have recorded from neurons in the medial premotor areas while monkeys made self-paced movements. It was found that some neurons in this region in fact fired as early as 0.6–2.6 seconds before movement onset. From the reported results it was also clear that this pattern of early firing for these neurons was consistent across trials. One other recent study has reported that the spatial pattern of fMRI activity from this region, at up to 5 seconds before action, can statistically predict the timing of action above chance level (Soon, Brass, Heinze, & Haynes, 2008).

Others have argued that the readiness potential may not reflect the specific and causal aspects of motor initiation. However, as mentioned earlier, it is likely that the readiness potential partly originates from the medial premotor areas. Lesion to these areas can abolish the production of spontaneous actions (Thaler, Chen, Nixon, Stern, & Passingham, 1995). These areas also contain neurons that code specific action plans (Shima & Tanji, 1998; Tanji & Shima, 1996). Further, when people use the Libet clock paradigm to time their own intentions, there is attentional modulation of activity in the medial pre-SMA (Lau, Rogers, Haggard, & Passingham, 2004), as if people were reading information off the area that is likely to be a source of the readiness potential.

The Libet clock method has also received considerable criticism. It involves timing across modalities and could be susceptible to various biases (Libet, 1985; Gomes, 2002; Joordens, van Duijn, & Spalek, 2002; Klein, 2002; Trevena & Miller, 2002). However, it is unlikely that all these biases

are in the direction that would help to narrow the gap between the onsets of the readiness potential and intention. Some have actually suggested that the different biases may point to different directions and thus just cancel each other out (Klein 2002). Also, in the original experiments by Libet and colleagues, there were control conditions that tested for the basic accuracy of the clock. They asked subjects to use the clock to time either the onset of movement execution, or in another condition to time the onset of tactile stimuli presented externally by the experimenter. Since the actual onsets of these events are objectively measurable, they could estimate the subjective error of onset reports produced by the clock method. They found the error to be on the order of about 50 ms; for example, people misestimate the time of action execution to be 50 ms earlier than it actually is. This size of error is considerably smaller than the gap between the onsets of the readiness potential and intention.

The basic results of Libet and colleagues have also been replicated in several different laboratories (e.g., Lau et al., 2004; Haggard & Eimer, 1999; Soon et al., 2008). In general, the same pattern is found, that the onset of intention is either around or later than 250 ms before action execution, which seems to confirm our intuition that conscious intentions seem to be followed by motor actions almost immediately. In fact, given that the readiness potential starts as early as 1–2 seconds before action execution, it is hard to imagine how the onset of intention could coincide with or precede the readiness potential, unless one thinks of intention as a kind of prior intention (Searle, 1983), like the general plan that is formed at the beginning of the experimental session when the subject agrees to produce some actions in the next half hour or so. We shall discuss this kind of higher-cognitive "intention" later in the chapter. However, the intention we are concerned with here is the immediate "urge" to produce the motor action (Libet, Wright, & Gleason, 1982).

Taken together, the evidence suggests that conscious intention—that is, the immediate feeling of motor initiation—is unlikely to be the "first unmoved mover" in triggering spontaneous motor movements. It is likely to be preceded by unconscious brain activity that may contribute to action initiation. What, then, is conscious intention for?

Conscious veto?

Libet's interpretation of the timing-of-intention results is that although intention may not be early enough to be the first cause of action, the fact that it occurs before action execution means that it could still be part of the causal chain. Maybe the decision to move is initiated unconsciously, but the awareness of intention may allow us to "veto"—that is, to cancel—the action.

This seems to be a possibility. Libet, Wright, and Gleason (1983) as well as other researchers (Brass & Haggard, 2007) have performed experiments where subjects prepare for an action and then cancel it at the last moment, just before it is executed. The fact that we have the ability to "veto" an action seems beyond doubt. The question, however, is whether having the conscious intention is critical. Can the choice of veto be preceded by unconscious activity, just as the intention to act is preceded by the readiness potential? Or maybe sometimes actions are unconsciously vetoed, even without our awareness?

Some recent evidence suggests that the conscious intention may not facilitate a veto. As mentioned earlier, when people were using the Libet clock to time the onset of their intentions, there was attentional modulation of activity in the pre-SMA (Lau et al., 2004). These data have been subsequently further analyzed (Lau, Rogers, & Passingham, 2006), and it has been shown that subjects who showed a large degree of attentional modulation also tended to report the onset of intention to be early. One interpretation could be that attention biases the judgment of onset to be earlier. It was found in another experiment that this was also true when people used the Libet clock to time the onset of the motor execution. The higher the level of fMRI activity modulated by attention, the earlier subjects reported the onset to be, even though on average subjects reported the onsets to be earlier than they actually were, a result which means that a bias to the negative (i.e., early) direction produced more erroneous rather than more precise reports. In general, the principle of attentional prior entry (Shore, Spence, & Klein, 2001) suggests that attention to an event speeds up its perception and negatively biases the reported onset. If this were true in the case of the Libet experiments, it could mean that attention might have exaggerated the 250-ms onset; that is, had subjects not been required to attend to their intentions in order to perform the timing tasks, the true onset of conscious intention might well have been much later than 250 ms prior to action execution. This possibility calls into question whether we have enough time to consider the veto.

Another study reported that some patients with lesion to the parietal cortex reported the onset of intention to be late as 50 ms prior to action execution (Sirigu et al., 2004). If the awareness of intention allows one to veto actions, one might expect these patients to have much less time to consciously evaluate spontaneous intentions and cancel the inappropriate ones. This could be quite disastrous to daily life functioning. Yet there were no such reports about these patients.

Finally, in another study (Lau, Rogers, & Passingham, 2007), single pulses of transcranial magnetic stimulation (TMS) were sent to the medial premotor areas (targeting the pre-SMA). Again, subjects were instructed to produce spontaneous movements and to time the onset of intentions and movement execution using the Libet clock. Surprisingly,

although TMS was applied *after* motor execution, it had an effect on the reported onsets. No matter whether TMS was applied immediately after action execution or with a 200-ms delay, the stimulation exaggerated the temporal distance between the reported onsets of intention and movement, as if people reported a prolonged period of conscious intending. One interpretation may be that TMS injected noisy activity into the area and the intention monitoring mechanism did not distinguish this from endogenously generated activity that is supposed to represent intention. However, what is crucial is the fact that the reported onsets can be manipulated even after the action is finished. This seems to suggest that our awareness of intention may be constructed after the facts, or at least not completely determined before the action is finished. If conscious intentions are not formed before the action, they certainly cannot play any role in facilitating veto, let alone causing it.

This interpretation may seem wild, but it is consistent with other proposals. For instance, on the basis of many ingenious experiments manipulating a subject's sense of agency, Wegner (2002) has suggested that the conscious will is an illusion. The sense of agency is often inferred post hoc, based on many contextual factors. Wegner cites experiments to support these claims. One example is a study on "facilitated communication" (Wegner, Fuller, & Sparrow, 2003). Subjects (playing the role of "facilitators") were asked to place their fingers on two keys of a keyboard, while a confederate (playing the role of "communicator") placed his or her fingers on top of those of the subject. Subjects were given headphones with which they listened to questions of varying difficulty. Confederates were given headphones as well, and subjects were led to believe that the confederates would be hearing the same questions, although in fact the confederates heard nothing. Subjects were told to detect subtle, unconscious movements in the confederate's fingers following each question. When such movements were detected, the subject was to press the corresponding key in order to answer on the confederate's behalf. It was found that subjects answered easy questions well above chance levels. If they had performed the task strictly according to the instructions, however, they should have performed at chance. Therefore, subjects must have been directing their own key presses. Nonetheless, they attributed a significant causal role for the key presses to the confederate. The degree to which subjects answered easy questions correctly was not correlated with the degree to which they attributed causal responsibility to confederates, suggesting that the generation of action and attribution of action to an agent are independent processes.

To summarize, although theorists have speculated that the awareness of intention may play some role in allowing us to cancel or edit our actions, considerable doubt has been cast by recent empirical evidence.

Exclusion and inhibition

Another kind of situation that seems to require conscious deliberation involves the need to avoid a particular action or response. This is related to "vetoing" as described previously, except that the action being inhibited is not necessarily self-paced and may be specified externally. One example would be to perform stem completion while avoiding a particular word. So, for instance, the experimenter may ask the subjects to produce any word starting with the letter *d* (i.e., completing a "stem") but avoid the word *dinner*. So subjects can produce *dog, danger, dear*, and so on, but if they produce the word *dinner*, it will be counted as an error. This is called the exclusion task (Jacoby, Lindsay, & Toth, 1992).

One interesting aspect of the exclusion task is that people can perform well only if they clearly see and remember the target of exclusion (i.e., the word *dinner* in the foregoing example). If the target of exclusion is presented very briefly and followed by a mask, such that it was only very weakly perceived, people may fail to exclude it (Debner & Jacoby, 1994; Merikle, Joordens, & Stolz, 1995). In fact, they tend to produce exactly the word they should be avoiding with higher likelihood than if they were not presented with the word at all. It has been argued that this exclusion failure phenomenon is the hallmark of unconscious processing (Jacoby et al., 1992). The weak perception of the target probably produced a representation for the word, but because the signal is not strong enough to reach the level of conscious processing, subjects are unable to inhibit the corresponding response.

In addition to the intuitive appeal, the notion that consciousness is required for exclusion is also supported by a case study of a blindsight patient (Persaud & Cowey, 2008). Subject GY has a lesion to the left primary visual cortex (V1) and reports that most of his right visual field is subjectively blind. However, in a forced-choice situation he can discriminate simple stimuli well above chance level in his "blind" field (Weiskrantz, 1986, 1997). In one study he was required to perform an exclusion task (Persaud & Cowey, 2008)—that is, to say the location (up or down) where the target was *not* presented. Whereas he could do this easily in the normal field, he failed the task when stimuli were presented to his blind field. Note that he was significantly worse than chance in the blind field, as if the unconscious signal drove the response directly and inflexibly, defying exclusion control. This result seems to support the account that consciousness is required for exclusion.

The general idea that inhibition requires consciousness seems to be supported by other studies too, including those that do not employ the exclusion paradigm. One study tested subjects' ability to ignore distracting moving dots while doing a central task that has nothing to do with the distracters (Tsushima, Sasaki, & Watanabe, 2006). It was

found that if the motion of the distracter was above the perceptual threshold, people could ignore the dots and inhibit the distraction successfully. Somewhat paradoxically, when the motion was below perceptual threshold, people could not ignore the dots and were distracted. The results from brain imaging seem to suggest that when the motion of the stimuli was strong, it activated the prefrontal cortex and triggered it to suppress the motion signal. When the motion of the stimuli was below perceptual threshold, however, the signal failed to trigger the inhibitory functions in the prefrontal cortex, and therefore the motion signal was not suppressed and thus remained distracting.

However, the notion that flexible control or inhibition of perceptual signal requires consciousness is not without its critics (Snodgrass 2002; Haase & Fisk, 2001; Visser & Merikle, 1999). One problem becomes clear when we consider the motion distracter example. "Conscious signal" here seems to be the same thing as a strong signal, driven by larger motion strength in the stimuli. Obviously, signals have to be strong enough to reach the prefrontal cortex in order to trigger the associating execution functions. Do unconscious stimuli fail to be excluded because we are not conscious of them, or is it just because the signal is not strong enough? Or, are the two explanations one and the same? Not all studies are subject to this argument. For instance, in the blindsight study (Persaud & Cowey, 2008), the subject failed to exclude in the blind field even when the contrast level would have given a performance that was similar to that in the normal visual field. So if we take forced-choice performance as an index of signal strength, the signal from the blind field was not weak in this sense. However, in most other cases we often take awareness to be the same as good performance. Are we justified in doing so? This is an important issue, and we will come back to it in the final section of the chapter.

Other researchers have reported evidence that seems to support unconscious inhibition. For instance, in one study (Snodgrass & Shevrin, 2006) people were asked to detect visually presented words. In certain conditions, some subjects showed detection performance that was significantly *worse* than chance. These words were presented so briefly that typically detection performance would be near chance. We usually take chance level as the objective threshold for conscious perception. Below-chance-level performance could be taken as evidence that the subjects did not consciously perceive the words. And yet, if they had no information at all regarding the words, performance should just be exactly at chance rather than below. It seems that these subjects were actively suppressing the words.

These are unusual cases and are somewhat hard to interpret. We take chance level as the objective threshold for conscious perception because when people perform at chance, it indicates that they do not have explicit information regarding the target of perception. However, if people perform significantly below chance, it means that somehow they have information regarding the detection, which violates the very logic we adopt to label perception unconscious. But in any case, the stimuli were supposed to be really weak, and it is intriguing that some subjects seem to be automatically suppressing the words. Are we to take these somewhat unusual cases as evidence to reject the notion that exclusion or inhibition requires consciousness? It seems that, logically, if we claim that a certain function *requires* consciousness, we should predict there will never be a case where one could perform such a function unconsciously. How seriously are we to take this logic and reject functions as requiring consciousness by a single experiment? We will return to this argument in the last section of the chapter.

Top-down cognitive control

So far we have discussed acts of volition that are relatively simple, like starting a motor movement or avoiding a particular action. Sometimes we also voluntarily prepare for a set of rules or action plans in order to satisfy a more abstract goal in mind. For instance, a telephone ring may usually trigger a particular action, for example, picking up the phone. However, when one visits friends at their homes, one may deliberately change the mapping between the stimulus (telephone ring) and action; that is, it would be more appropriate to sit still or ask the host to pick up the phone, rather than pick it up oneself. This volitional change of stimulus-response contingency is an example of top-down cognitive control.

It has been suggested that top-down cognitive control may require consciousness (Dehaene & Naccache, 2001). The idea is that unconscious stimuli can trigger certain prepared actions, as demonstrated in studies in subliminal priming (Kouider & Dehaene, 2007). However, the preparation or setting up of the stimulus-response contingency may require consciousness.

However, recent studies suggest that this conclusion might not be true, in the sense that unconscious information seems to be able to influence or even trigger top-down cognitive control too (Mattler, 2003; Lau & Passingham, 2007). In one study subjects had to prepare to do a phonological or semantic judgment, based on the orientation of a figure they saw (figure 83.3). In every trial, if they saw a square, they had to prepare to judge whether an upcoming word has two syllables (e.g., *table*) or not (e.g., *milk*). If they saw a diamond, they had to prepare to judge whether an upcoming word refers to a concrete object (e.g., *chair*) or an abstract idea (e.g., *love*). In other words, they had to perform top-down cognitive control based on the instruction figure (square or diamond). However, before the instruction figure was presented, there was actually an invisible prime figure, which could also be a diamond or a square. It was found that the prime could

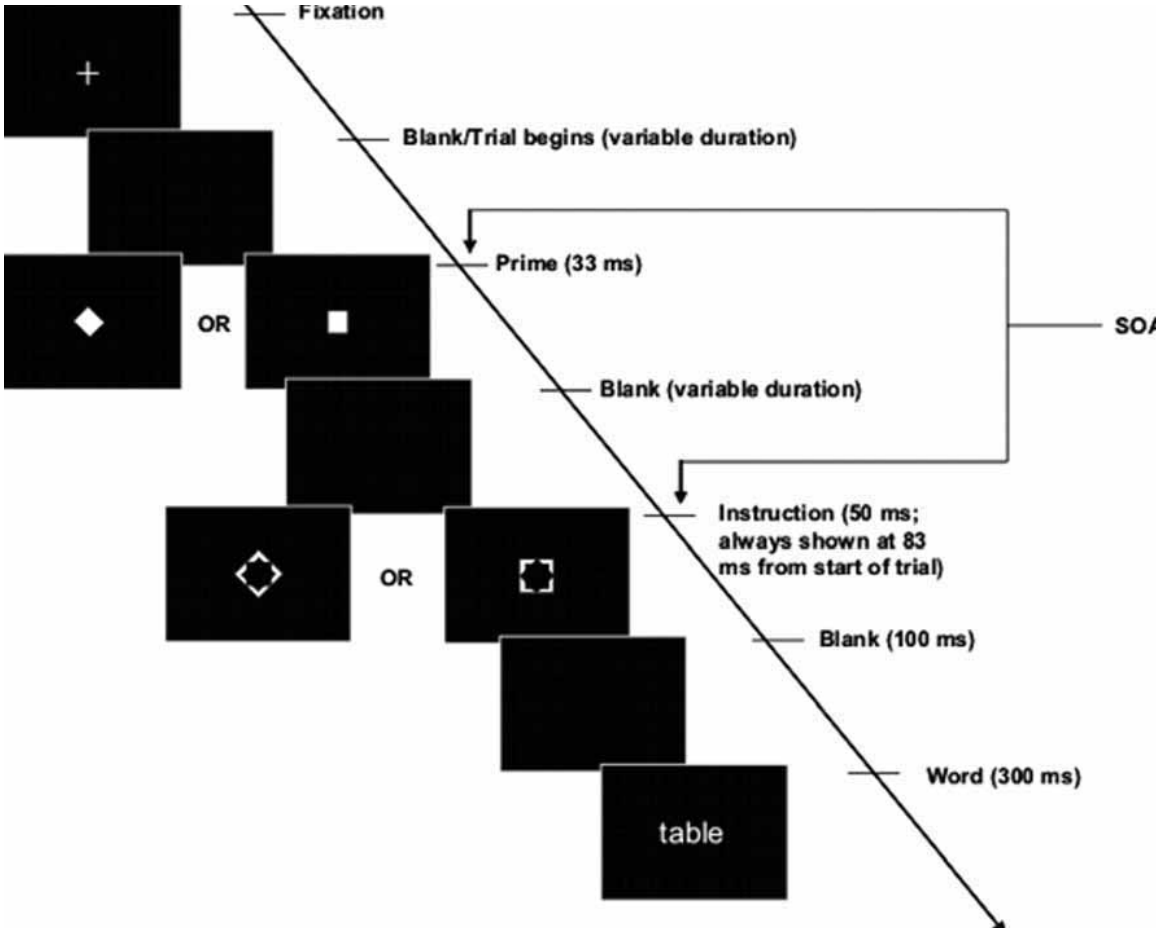


FIGURE 83.3 Experimental paradigm of Lau and Passingham (2007). Subjects view briefly presented words and perform either a phonological task (is the word one syllable or two syllables?) or a semantic task (does the word name something concrete or abstract?). Before word presentation, subjects are instructed which task to perform on a given trial by a visual symbol (a square for the phonological task or a diamond for the semantic task). The symbolic instruction itself acts as a metacontrast mask for an earlier prime, also a square or a diamond. Because the prime is briefly presented and masked, it is not consciously perceived. On half of the trials, the prime is congruent with the instruction, and on the other half, incongruent. Behavioral and imaging results suggest that the unconscious primes affected top-down task switching. When primes were incongruent with instructions, accuracy fell, reaction time increased, and brain regions corresponding to the task indicated by the prime were partially activated (all relative to the prime-congruent condition). But when the stimulus onset asynchrony (SOA) between prime and instruction was lowered, such that primes became visible, the priming effect was not evident. This double dissociation suggests that the interference of incongruent primes on task switching cannot be attributed to conscious processing. (Adapted from Lau & Passingham, 2007.)

impair subjects' performance when it suggested the alternative (i.e., wrong) task to the subjects (incongruent condition). One could argue that this result occurred only because the prime distracted the subjects on a perceptual level and did not really trigger cognitive control. However, the experiment was performed in the fMRI scanner, and the brain recordings suggest that when being primed to perform the wrong task, subjects used more of the wrong neural resources too (Lau & Passingham, 2007). That is, areas that are more sensitive to phonological or semantic processing showed increased activity when the explicit instruction figure made subjects perform the phonological and semantic tasks, respectively. The invisible primes also seem to be able to trigger activations in task-sensitive areas. This result seems to suggest that they can influence or exercise top-down cognitive control.

Another study examines how unconscious information affects our high-level objectives by focusing on how the potential reward influences our level of motivation (Pessiglione et al., 2007). Subjects squeezed a device to win a certain amount of money. The harder they squeezed, the more money they would win. However, the size of the stake in question for a particular trial was announced in the beginning by presenting the photo of a coin. The coin could either be a British pound (~2 U.S. dollars) or a penny (~2 U.S. cents), and it signified the monetary value of the maximal reward for that trial. Not surprisingly, people squeezed harder when the stakes were high, but interestingly, the same

pattern of behavior was observed even when the figure of the coin was masked such that subjects reported not seeing it. This finding suggests that unconscious information can influence our level of motivation as well.

If unconscious information alone is sufficient to exercise all these sophisticated top-down control functions, why do we need to be conscious at all?

How can we find the true function of consciousness?

The foregoing is not meant to be an exhaustive review of all studies on the potential functions of consciousness. We selected some examples from a few areas that are particularly related to volition and discussed the role that consciousness may play. There may, of course, be other psychological functions that require consciousness.

Yet one cannot help feeling that there seems to be some inherent limitation to this whole enterprise of research. If we claim that a certain function requires consciousness, strictly speaking, the interpretation could be that the function should never be able to be performed unconsciously. Of course, one could make the weaker claim that a certain function is usually or most suitably performed consciously, and when consciousness fails, unconscious processing can act as a backup. This is similar to arguing that one function of having legs is to facilitate locomotion; if we lose our legs, we could still move around, albeit poorly. However, let us assume that one is to make the stronger prediction that such functions should never be able to be performed unconsciously. In principle, it would only take a single experiment to falsify that. This assumption explains why this review may seem biased in that we focus on studies that show the power of the unconscious, rather than studies demonstrating functions that definitely require consciousness. In principle, falsifying the claim that a certain function requires consciousness is straightforward. But this is not the case for demonstrating functions that would always require consciousness.

One can, of course, try to show that subjects could normally do a task if the relevant information is consciously perceived. And then one tries to "knock out" the conscious perception for such information and show that the task could no longer be performed, or that it is performed at an additional cost, that is, slower or with more errors. But how would one know that in "knocking out" the conscious perception, one does not "knock out" too much? One typically suppresses conscious perception by visual masking, by using brief presentation, by distracting the subject, by applying transcranial magnetic stimulation, by pharmacological manipulations, and so on. But all of these could potentially impair the unconscious as well as the conscious signal. Could it be that in cases where the perception has been rendered unconscious, the signal is just no longer strong enough to drive the function in question? This intepretation would mean that, in principle, it would be possible for a future study to find the optimal procedure or setup to just render the information unconscious, without reducing the signal strength too much. And in that case the subjects may be able to perform the task in question. That result would falsify our claim.

Consequently, in looking for functions that require consciousness, we need to adopt some different strategies. One potentially useful approach is to try to demonstrate something akin to a "double dissociation." When conscious perception is suppressed, we often find that a sophisticated function (e.g., top-down cognitive control) can no longer be performed, though some simpler function (e.g., priming for a prepared motor response) may still be activated by unconscious information. From the foregoing discussion, one could see that this outcome may not be as surprising or informative as it seems. It could be just that the unconscious signal is too weak to drive the relatively sophisticated function. A demonstration of the opposite would, however, be much more convincing: If after suppression of conscious perception, the subjects can still perform a rather sophisticated function but fail to perform a simple function, that finding would suggest that the simple function really requires consciousness. In this case, it could not be that the suppression of conscious perception has taken away too much of the signal strength, because if that were the case then the subjects should not be able to perform the relatively sophisticated function (figure 83.4). Understanding this "double dissociation"

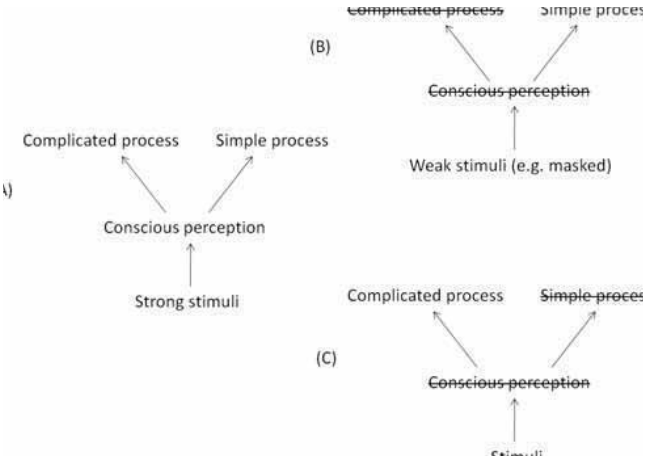


Figure 83.4 (*A*) The normal situation for conscious perception. Stimuli are strong enough to drive processes of different complexity. (*B*) A typical situation for unconscious perception. Stimuli are weak such that complicated processes are no longer activated, though simple processes can still be triggered. It could be argued that this result is not surprising, since we may expect that complicated processes require a stronger signal. (*C*) A potentially more informative situation. If one could find a stimulus that is not consciously perceived, yet is sufficiently strong to trigger a complicated process, then the relatively simple process that the stimulus does not drive would seem to critically depend on consciousness.

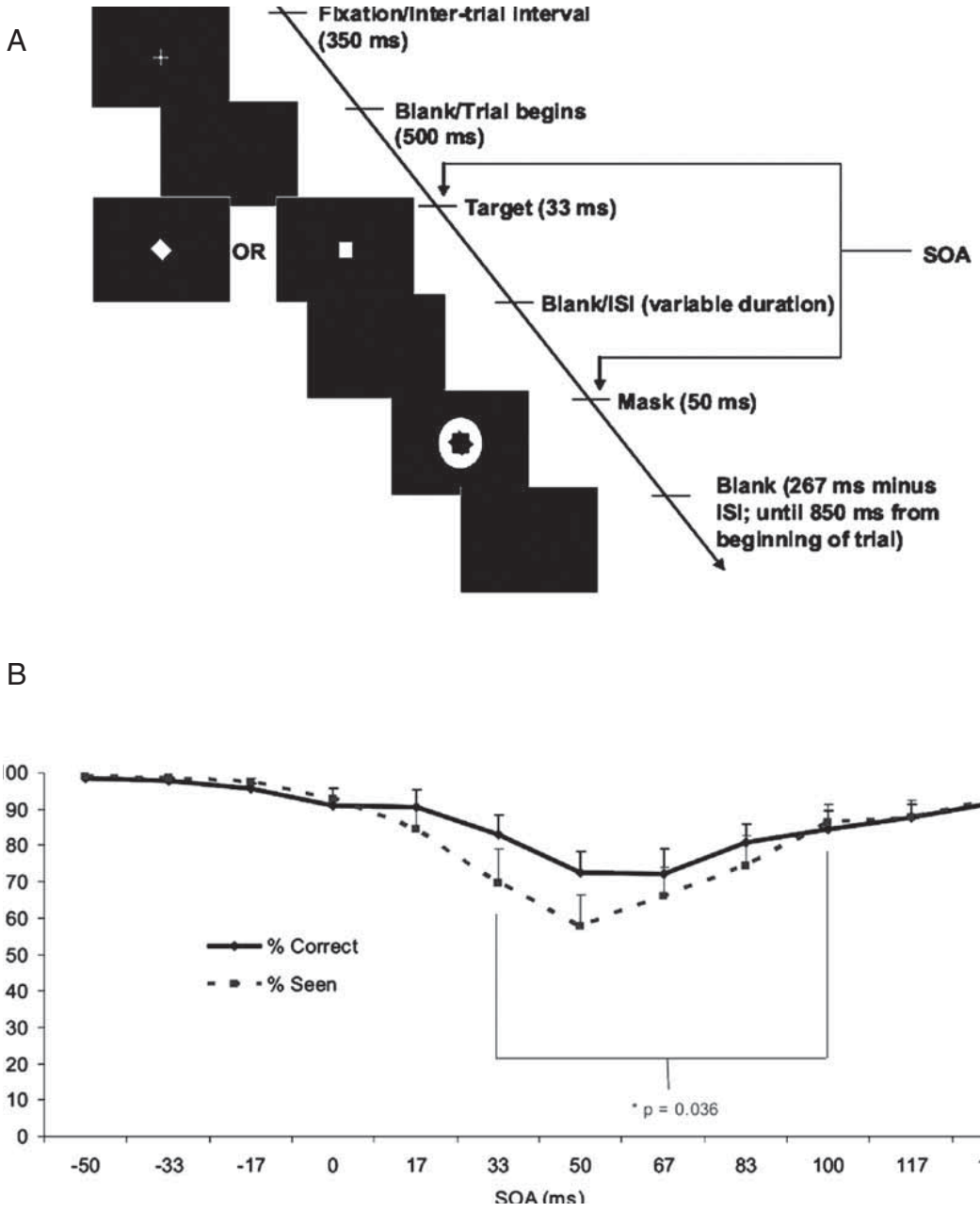


FIGURE 83.5 Inducing "relative blindsight" in normal observers using metacontrast masking. (*A*) Metacontrast masking paradigm. The subject is presented with a visual target (in this case, either a square or diamond). Afterward, a metacontrast mask is presented. The mask differentially affects discrimination accuracy and visual awareness of the target as a function of stimulus onset asynchrony (SOA). (*B*) Discrimination accuracy and visual awareness as a function of metacontrast mask SOA. The metacontrast mask creates a characteristic U-shaped function of performance versus SOA. At shorter and longer SOAs, discrimination accuracy is high, but it dips at intermediate SOAs. The same is true for visual awareness, but the shape of the awareness-masking function is not perfectly symmetrical with respect to the performance-masking function. That is, there are certain SOAs at which forced-choice performance is matched but visual awareness differs significantly (e.g., as illustrated in the SOAs of 33 ms and 100 ms). Such performance-matched conditions could be used to investigate the functions of consciousness. If some task can be performed better in the condition of higher subjective visibility, it can plausibly be said to require visual awareness. Because forced-choice discrimination accuracy is matched across the two conditions, the superior performance of the task in the high-visibility condition cannot be attributed to a difference in signal strength. (Adapted from Lau & Passingham, 2006.)

approach helps us see the logic behind how we could deal with signal strength as a confounding variable. However, one problem is that it is unclear what is the most convincing way to define "sophisticated/complicated" functions versus "simple" functions.

An alternative approach may be to directly match for signal strength between the conscious and the unconscious conditions. This might seem difficult because conscious signals may seem to be strong in general. However, as discussed earlier, blindsight subjects can perform forced-choice discrimination on visual stimuli well above chance, even when they claim that conscious awareness is missing. Forced-choice performance is often taken as an objective estimate of signal strength; the detection theoretical measure d' is mathematically just the signal-to-noise ratio. In blindsight subject GY, where only half of the visual field lacks awareness, we can imagine presenting weak stimuli to the normal visual field such that forced-choice performance would match that in the blind field (Weiskrantz, Barbur, & Sahraie, 1995). This way we can test whether certain functions cannot be performed based on information presented to the blind field, a procedure which may shed light on when consciousness is required.

One may argue that blindsight patients are rare and that the way their brains process visual information may not generalize to intact brains. However, there are other paradigms where in normal subjects one could match for forced-choice performance and yet produce a difference in the level of conscious awareness. For instance, in one study (Lau & Passingham, 2006) metacontrast masking was used to create similar conditions where forced-choice discrimination accuracy for the visual targets was matched, and yet the subjective reports of how often subjects saw the identity of the targets differed (figure 83.5). One could imagine presenting these stimuli to subjects and seeing whether they drive a certain function with different effectiveness. If the subjects perform better in the condition where subjective conscious awareness of the stimuli is more frequent, one could argue that this function is likely to depend critically on consciousness.

Conclusion

Acts of volition are accompanied by a sense of conscious effort or intention. The fact that we feel the conscious effort is not in doubt. What is less clear is whether the processes underlying the conscious experience directly contribute to the execution of the actions, in a way that is not accomplished by unconscious processes just as effectively. The general picture seems to be that many sophisticated functions can be performed unconsciously or driven by unconscious information.

Does this conclusion mean that consciousness has no special function at all? The answer is not yet clear. It is likely that some psychological functions require consciousness. That is, there may be some functions that can only be performed poorly with unconscious information. Or, there may even be functions that can never be performed unconsciously. But experiments have not yet been able to convincingly pin them down.

Future research will have to overcome the following problem. If we assume that conscious perception is always accompanied by stronger and longer-lasting signals that are more effective than unconscious signals in propagating themselves throughout the brain, then consciousness would certainly be associated with the functions of these strong signals. However, in studies of blindsight (Weiskrantz et al., 1995), as well as in normals (Lau & Passingham, 2006), it has been shown that signal strength as indicated by forced-choice performance is not always one and the same as conscious awareness. Therefore, future studies may need to focus on identifying the functions that really cannot be performed unconsciously, even when the signal strength is sufficiently strong. This approach may help to reveal the true function of consciousness.

ACKNOWLEDGMENTS The author thanks David Rosenthal and Uriah Kriegel for comments.

REFERENCES

Ball, T., Schreiber, A., Feige, B., Wagner, M., Lücking, C. H., & Kristeva-Feige, R. (1999). The role of higher-order motor areas in voluntary movement as revealed by high-resolution EEG and fMRI. *NeuroImage*, *10*(6), 682–694.

Brass, M., & Haggard, P. (2007). To do or not to do: The neural signature of self-control. *J. Neurosci.*, *27*(34), 9141–9145.

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2003). The preparation and readiness for voluntary movement: A high-field event-related fMRI study of the Bereitschafts-BOLD response. *NeuroImage*, *20*(1), 404–412.

Debner, J. A., & Jacoby, L. L. (1994). Unconscious perception: Attention, awareness, and control. *J. Exp. Psychol. Learn. Mem. Cogn.*, *20*(2), 304–317.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37.

Erdler, M., Beisteiner, R., Mayer, D., Kaindl, T., Edward, V., Windischberger, C., et al. (2000). Supplementary motor area activation preceding voluntary movement is detectable with a whole-scalp magnetoencephalography system. *NeuroImage*, *11*(6), 697–707.

Gomes, G. (2002). The interpretation of Libet's results on the timing of conscious events: A commentary. *Conscious. Cogn.*, *11*(2), 221–230; discussion, 308–313, 314–325.

Haase, S. J., & Fisk, G. (2001). Confidence in word detection predicts word identification: Implications for an unconscious perception paradigm. *Am. J. Psychol.*, *114*(3), 439–468.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Exp. Brain Res.*, *126*(1), 128–133.

Jacoby, L. L., Lindsay, D. S., & Toth, J. P. (1992). Unconscious influences revealed: Attention, awareness, and control. *Am. Psychol.*, *47*(6), 802–809.

Joordens, S., van Duijn, M., & Spalek, T. M. (2002). When timing the mind one should also mind the timing: Biases in the measurement of voluntary actions. *Conscious. Cogn.*, *11*(2), 231–240; discussion, 308–313.

Klein, S. (2002). Libet's research on the timing of conscious intention to act: A commentary. *Conscious. Cogn.*, *11*(2), 273–279; discussion, 304–325.

Kornhuber, H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkurbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Arch.*, *284*, 1–17.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *362*(1481), 857–875.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl. Acad. Sci. USA*, *103*(49), 18763–18768.

Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *J. Neurosci.*, *27*(21), 5805–5811.

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. *Science*, *303*(5661), 1208–1210.

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2006). On measuring the perceived onsets of spontaneous actions. *J. Neurosci.*, *26*(27), 7265–7271.

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2007). Manipulating the experienced onset of intention after action execution. *J. Cogn. Neurosci.*, *19*(1), 81–90.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav. Brain Sci.*, *8*, 529–566.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*, *106*(Pt. 3), 623–642.

Libet, B., Wright, E. W., & Gleason, C. A. (1982). Readiness-potentials preceding unrestricted "spontaneous" vs. pre-planned voluntary acts. *Electroencephalogr. Clin. Neurophysiol.*, *54*(3), 322–335.

Libet, B., Wright, E. W., & Gleason, C. A. (1983). Preparation- or intention-to-act, in relation to pre-event potentials recorded at the vertex. *Electroencephalogr. Clin. Neurophysiol.*, *56*(4), 367–372.

Mattler, U. (2003). Priming of mental operations by masked stimuli. *Percept. Psychophys.*, *65*(2), 167–187.

Merikle, P. M., Joordens, S., & Stolz, J. A. (1995). Measuring the relative magnitude of unconscious influences. *Conscious. Cogn.*, *4*(4), 422–439.

Miller, J., & Trevena, J. A. (2002). Cortical movement preparation and conscious decisions: Averaging artifacts and timing biases. *Consciousness Cogn.*, *11*(2), 308–313.

Persaud, N., & Cowey, A. (2008). Blindsight is unlike normal conscious vision: Evidence from an exclusion task. *Consciousness Cogn.*, *17*(3), 1050–1055.

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., et al. (2007). How the brain translates money into force: A neuroimaging study of subliminal motivation. *Science*, *316*(5826), 904–906.

Romo, R., & Schultz, W. (1987). Neuronal activity preceding self-initiated or externally timed arm movements in area 6 of monkey cortex. *Exp. Brain Res.*, *67*(3), 656–662.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind* (p. 278). Cambridge, UK: Cambridge University Press.

Shima, K., & Tanji, J. (1998). Both supplementary and presupplementary motor areas are crucial for the temporal organization of multiple movements. *J. Neurophysiol.*, *80*(6), 3247–3260.

Shore, D. I., Spence, C., & Klein, R. M. (2001). Visual prior entry. *Psychol. Sci.*, *12*(3), 205–212.

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., et al. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nat. Neurosci.*, *7*(1), 80–84.

Snodgrass, M. (2002). Disambiguating conscious and unconscious influences: Do exclusion paradigms demonstrate unconscious perception? *Am. J. Psychol.*, *115*(4), 545–579.

Snodgrass, M., & Shevrin, H. (2006). Unconscious inhibition and facilitation at the objective detection threshold: Replicable and qualitatively different unconscious perceptual effects. *Cognition*, *101*(1), 43–79.

Soon, C. S., Brass, M., Heinze, H., & Haynes, J. (2008). Unconscious determinants of free decisions in the human brain. *Nat. Neurosci.*, *11*(5), 543–545.

Tanji, J., & Shima, K. (1996). Supplementary motor cortex in organization of movement. *Eur. Neurol.*, *36*(Suppl. 1), 13–19.

Thaler, D., Chen, Y. C., Nixon, P. D., Stern, C. E., & Passingham, R. E. (1995). The functions of the medial premotor cortex. I. Simple learned movements. *Exp. Brain Res.*, *102*(3), 445–460.

Trevena, J. A., & Miller, J. (2002). Cortical movement preparation before and after a conscious decision to move. *Consciousness Cogn.*, *11*(2), 162–190; discussion, 314–325.

Tsushima, Y., Sasaki, Y., & Watanabe, T. (2006). Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science*, *314*(5806), 1786–1788.

Visser, T. A., & Merikle, P. M. (1999). Conscious and unconscious processes: The effects of motivation. *Consciousness Cogn.*, *8*(1), 94–113.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Wegner, D. M., Fuller, V. A., & Sparrow, B. (2003). Clever hands: Uncontrolled intelligence in facilitated communication. *J. Pers. Soc. Psychol.*, *85*(1), 5–19.

Weilke, F., Spiegel, S., Boecker, H., von Einsiedel, H. G., Conrad, B., Schwaiger, M., et al. (2001). Time-resolved fMRI of activation patterns in M1 and SMA during complex voluntary movement. *J. Neurophysiol.*, *85*(5), 1858–1863.

Weiskrantz, L. (1986). *Blindsight: A case study and implications* (p. 187). Oxford, UK: Oxford University Press.

Weiskrantz, L. (1997). *Consciousness lost and found: A neuropsychological exploration* (1st ed., p. 304). New York: Oxford University Press.

Weiskrantz, L., Barbur, J. L., & Sahraie, A. (1995). Parameters affecting conscious versus unconscious visual discrimination with damage to the visual cortex (V1). *Proc. Natl. Acad. Sci. USA*, *92*(13), 6122–6126.

84 Toward a Theory of Consciousness

GIULIO TONONI AND DAVID BALDUZZI

ABSTRACT Cognitive neuroscience provides us with both clues and paradoxes about the neural substrate of consciousness. For example, we know that certain corticothalamic circuits are essential for conscious experience, whereas cerebellar circuits are not, despite their huge numbers. We also know that consciousness wanes during slow-wave sleep and generalized seizures, despite levels of neural activity that are comparable to wakefulness. To understand why this is so, empirical observations must be related to a theory that says, in a principled manner, what consciousness is and how it can be generated. This chapter introduces the integrated information theory. Starting from phenomenology and making a critical use of thought experiments, the theory claims that consciousness *is* integrated information. Specifically, (1) the quantity of consciousness is given by the amount of integrated information generated by a complex of elements, and (2) the quality of experience, such as the "redness" of red, is given by the set of informational relationships within that complex. Integrated information (symbol, Φ) is defined as the amount of information generated by causal interactions within a complex of elements, above and beyond the information generated independently by its parts. Qualia space (symbol, Q) is a space where each axis represents a possible state of the complex, each point is a probability distribution of its states, and arrows between points represent the informational relationships generated by causal interactions among its elements. Together, the set of informational relationships within a complex specifies a shape in Q that in turn specifies a particular experience. Several observations concerning the neural substrate of consciousness fall naturally into place within the integrated information framework.

Consciousness poses two related problems (Tononi, 2001). The first is to understand what features of the brain determine the extent to which consciousness is present. For example, why are certain corticothalamic circuits important for conscious experience, whereas cerebellar circuits are not, though the number of neurons in the two structures is comparable and their neurobiological organization is similarly complicated? And why is consciousness strikingly reduced during deep slow-wave sleep or during absence seizures, despite high levels of neuronal firing?

The second problem of consciousness is to understand what features of the brain determine the specific way consciousness is experienced—what is responsible for, say, the "redness" of red? We know that the activity of specific cortical areas contributes specific dimensions of conscious experience—auditory cortex to sound, visual cortex to shapes and colors. Why is this so? Solving the first problem means that we would know to what extent a physical system can generate consciousness—the *quantity* or level of consciousness. Solving the second problem means that we would know what kind of consciousness it generates—the *quality* or content of consciousness.

Earlier chapters in part X have reviewed empirical evidence on the neural correlates of consciousness (Crick & Koch, 2003; Koch, 2004). Here we focus instead on the theoretical foundations of consciousness science. Specifically, we discuss the integrated information theory of consciousness (Tononi, 2004), according to which brain mechanisms generate experience to the extent that they generate integrated information. In what follows, we first consider phenomenological thought experiments indicating that subjective experience has to do with the generation of integrated information; we then consider ways of defining and measuring integrated information; next, we show how basic facts about consciousness and the brain can be accounted for in terms of integrated information; finally, we examine some ideas about how the second problem of consciousness can be addressed.

The first problem: What determines the quantity of consciousness?

Everyday experience indicates that consciousness has a physical substrate and that the physical substrate must be working in the proper way for us to be fully conscious—it is enough to fall asleep, receive a blow on the head, or take certain drugs such as anesthetics, to affect our consciousness dramatically. These observations raise the question of what are the necessary and sufficient conditions that determine whether consciousness is present and to what extent. The question is a general one with multiple implications. For example, is a person with akinetic mutism—awake with eyes open, but mute, immobile, and unresponsive—conscious or not? How much consciousness is there during sleepwalking or psychomotor seizures? Are newborn babies conscious,

GIULIO TONONI and DAVID BALDUZZI Department of Psychiatry, University of Wisconsin, Madison, Wisconsin

and to what extent? Are animals conscious? If so, are some animals more conscious than others? Could we one day build conscious artifacts using nonneural ingredients (Koch & Tononi, 2008), or at least augment our own consciousness with implanted chips?

Consciousness as Integrated Information The *integrated information theory (IIT)* of consciousness claims that, at the fundamental level, consciousness *is* integrated information, and that its quality is given by the informational relationships generated by a complex of elements. These claims stem from realizing that information and integration are the essential properties of our own experience. This fact may not be immediately evident, perhaps because, being endowed with consciousness most of the time, we tend to take its gifts for granted. To regain some perspective, it is useful to resort to two thought experiments, one involving a photodiode and the other a digital camera.

Information Consider the following: You are facing a blank screen that is alternately on and off, and you have been instructed to say "light" when the screen turns on and "dark" when it turns off. A photodiode—a simple light-sensitive device—has also been placed in front of the screen. It contains a sensor that responds to light with an increase in current and a detector connected to the sensor that says "light" if the current is above a certain threshold and "dark" otherwise. The first problem of consciousness reduces to this: when you distinguish between the screen being on or off, you have the subjective experience of seeing light or dark. The photodiode can also distinguish between the screen being on or off, but presumably it does not have a subjective experience of light and dark. What is the key difference between you and the photodiode?

According to the IIT, the difference has to do with how much information is generated when that distinction is made. Information is classically defined as reduction of uncertainty: the more numerous the alternatives that are ruled out, the greater the reduction of uncertainty, and thus the information. It is usually measured using the entropy function, which is the logarithm of the number of alternatives (assuming they are equally likely). For example, tossing a fair coin and obtaining heads corresponds to $\log_2(2) = 1$ bit of information, because there are just two alternatives; throwing a fair die yields $\log_2(6) = 2.59$ bits of information, because there are six alternatives.

Let us now compare the photodiode with you. When the blank screen turns on, the mechanism in the photodiode tells the detector that the current from the sensor is above rather than below the threshold, so it reports "light." In performing this discrimination between two alternatives, the detector in the photodiode generates $\log_2(2) = 1$ bit of information. When you see the blank screen turn on, however, the situation is quite different. Though you may think you are performing the same discrimination between light and dark as the photodiode, you are in fact discriminating among a much larger number of alternatives, thereby generating many more bits of information.

This point is easy to see. Just imagine that, instead of turning light and dark, the screen were to turn red, then green, then blue, and then display, one after the other, every frame from every movie that was ever produced. The photodiode, inevitably, would go on signaling whether the amount of light for each frame is above or below its threshold: to a photodiode, things can only be one of two ways, so when it reports "light," it really means just "this way" versus "that way." For you, however, a light screen is different not only from a dark screen, but also from a multitude of other images, so when you say "light," it really means this specific way versus countless other ways, such as a red screen, a green screen, a blue screen, this movie frame, that movie frame, and so on for every movie frame (not to mention for any sound, smell, thought, or any combination of the above). Clearly, each frame looks different to you, implying that some mechanism in your brain must be able to tell it apart from all the others. So when you say "light," whether you think about it or not (and you typically will not), you have just made a discrimination among a very large number of alternatives, and thereby generated many bits of information.

This point is so deceivingly simple that it is useful to elaborate a bit on why, although a photodiode may be as good as we are in detecting light, it cannot possibly see light the way we do—in fact, it cannot possibly "see" anything at all. Hopefully, by realizing what the photodiode lacks, we may appreciate what allows us to consciously "see" the light.

The key is to realize how the many discriminations that we can do, and that the photodiode cannot, affect the *meaning* of the discrimination at hand, the one between light and dark. For example, the photodiode has no mechanism to discriminate colored from achromatic light, even less to tell which particular color the light might be. As a consequence, all light is the same to it, as long as it exceeds a certain threshold. So for the photodiode "light" cannot possibly mean achromatic as opposed to colored, not to mention any particular color. Also, the photodiode has no mechanism to distinguish between a homogeneous light and a bright shape—any bright shape—on a darker background. So for the photodiode light cannot possibly mean full field as opposed to a shape—any of countless particular shapes. Worse, the photodiode does not even know that it is detecting a visual attribute—the "visualness" of light—as it has no mechanism to tell visual attributes from nonvisual ones, such as sounds or smells, not to mention which particular sounds or smells, and so on. As far as it knows, the photodiode might

just as well be a thermistor—it has no way of knowing whether it is sensing light versus dark or hot versus cold.

In short, the only specification a photodiode can make is whether things are this or that way: any further specification is impossible because it does not have mechanisms for it. Therefore, when the photodiode detects "light," such "light" cannot possibly mean what it means for us—it does not even mean that it is a visual attribute. By contrast, when we see "light," we are implicitly being much more specific: we simultaneously specify that things are this way rather that that way (light as opposed to dark), that whatever we are discriminating is not colored (in any particular color), does not have a shape (any particular one), is visual as opposed to auditory or olfactory, sensory as opposed to thoughtlike, and so on. To us, then, light is much more meaningful precisely because we have mechanisms that can discriminate this particular state of affairs we call "light" against a large number of alternatives.

According to the IIT, it is all this added meaning, provided implicitly by *how* we discriminate pure light from all these alternatives, that increases its level of consciousness. This central point may be appreciated either by "subtraction" or by "addition." By subtraction, one may realize that our being conscious of "light" would degrade more and more—would lose its noncoloredness, its nonshapedness, would even lose its visualness—as its meaning is progressively stripped down to just "one of two ways," as with the photodiode. By addition, one may realize that we can only see "light" as we see it, as progressively more and more meaning is added by specifying how it differs from countless alternatives. Either way, the theory says that, the more specifically one's mechanisms discriminate between what pure light is and what it is not—the more they specify what light means—the more one is conscious of it.

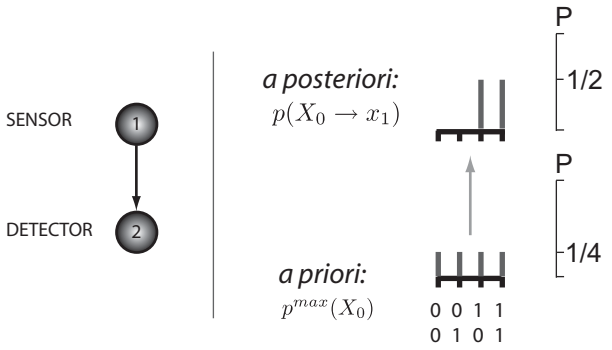
Integration Information—the ability to discriminate among a large number of alternatives—is thus essential for consciousness. However, information always implies a *point of view* or perspective (Metzinger, 2003), and we need to be careful about what that point of view might be (information for whom?) To see why, consider another thought experiment, this time involving a digital camera—say, one whose sensor chip is a collection of a million binary photodiodes, each sporting a sensor and a detector. Clearly, taken as a whole, the camera's detectors could distinguish among $2^{1,000,000}$ alternative states, an immense number, corresponding to 1 million bits of information. Indeed, the camera would easily respond differently to every frame from every movie that was ever produced. Yet few would argue that the camera is conscious. What is the key difference between you and the camera?

According to the IIT, the difference has to do with integrated information. From the point of view of an external observer, the camera may be considered as a single system with a repertoire of $2^{1,000,000}$ states. In reality, however, the chip is not an integrated entity: since its 1 million photodiodes have no way to interact, each photodiode performs its own local discrimination between a low and a high current, completely independent of what every other photodiode might be doing. In reality, the chip is just a collection of 1 million independent photodiodes, each with a repertoire of 2 discriminable states. In other words, there is no intrinsic point of view associated with the camera chip as a whole. This point is easy to see: if the sensor chip were cut into 1 million pieces each holding its individual photodiode, the performance of the camera would not change at all.

By contrast, you discriminate among a vast repertoire of states as an integrated system, one that cannot be broken down into independent components each with its own separate repertoire. Phenomenologically, every experience is an integrated whole, one that means what it means by virtue of being one and that is experienced from a single point of view. For example, the experience of a red square cannot be decomposed into the separate experience of red and the separate experience of a square. Similarly, experiencing the full visual field cannot be decomposed into experiencing separately the left half and the right half: such a possibility does not even make sense to us, since experience is always whole. Indeed, the only way to split an experience into independent experiences seems to be to split the brain in two, as in patients who underwent the section of the corpus callosum to treat severe epilepsy (Gazzaniga, 2005). Such patients do indeed experience the left half of the visual field independently of the right side, but then the surgery has created two separate consciousnesses instead of one. Mechanistically then, underlying the unity of experience must be causal interactions among certain elements within the brain. This means that these elements work together as an integrated system, and it is for this reason that their performance breaks down if they are disconnected, unlike the camera.

Measuring Integrated Information This phenomenological analysis suggests that to generate consciousness, a physical system must be able to discriminate among a large repertoire of states (information) *and* it must be unified; that is, it should do so as a single system, not decomposable into a collection of causally independent parts (integration). But how can one measure integrated information? As we shall explain, the central idea is to quantify the information generated by a system, above and beyond the information generated independently by its parts (Balduzzi & Tononi, 2008; Tononi, 2001, 2004).

Information First, we need to evaluate how much information is generated by the system. Consider the system of two binary units in figure 84.1, which can be thought of as an



$$ei(X_0 \to x_1) = H\left[p(X_0 \to x_1) \| \, p^{max}(X_0)\right] = 1 \text{ bit}$$

FIGURE 84.1 Effective information. A "photodiode" consisting of a sensor and detector unit. The detector unit turns on if the sensor's current is above a threshold. Here both units are on (binary 1, indicated in gray). For the entire system (sensor unit, detector unit) there are four possible states: (00,01,10,11). The a priori distribution $p^{max}(X_0) = (¼, ¼, ¼, ¼)$ is the maximum entropy distribution on the four states. Given that the detector is on, the mechanism specifies that the sensor must have been on, thus ruling out two of the four possible states (00,01). The a posteriori distribution is therefore $p(X_0 \to x_1) = (0, 0, ½, ½)$. Note that the prior state of the detector makes no difference to the current state of the system, so the states (10,11) are indistinguishable to the mechanism. Relative entropy (Kullback-Leibler divergence) between two probability distributions p and q is $H[p|q] = \Sigma_i \, p_i \log_2 (p_i / q_i)$ so that effective information (the entropy of the a posteriori relative to the a priori distributions) associated with output $x_1 = 11$ is 1 bit.

idealized version of a photodiode composed of a sensor *S* and a detector *D*. The system is characterized by a state it is in, which in this case is 11, and by a mechanism. This is mediated by a connection (arrow) between the sensor and the detector that implements a causal interaction: in this case, the elementary mechanism of the system is that the detector checks the state of the sensor and turns on if the sensor was on, and off otherwise (more generally, the specific causal interaction can be described by an input-output table).

A priori, a system of two binary elements could be in any of four possible states (00,01,10,11) with equal probability: $p = (¼, ¼, ¼, ¼)$. Formally, this a priori (potential) repertoire is represented by the maximum entropy or uniform distribution of possible system states, which expresses complete uncertainty. However, given the mechanism of the system and the state it is in (in this case $x_1 = 11$), uncertainty is reduced—namely, the uncertainty about the previous state of the system. A posteriori, a system with this mechanism being in state 11 *specifies* that the previous system state x_0 must have been either 11 or 10, rather than 00 or 01, corresponding to $p = (0, 0, ½, ½)$ (in this system, there is no mechanism to specify the detector state, which remains uncertain). Formally, then, the mechanism and the state 11 specify an a posteriori distribution or repertoire of system states that could have caused (led to) x_1, while ruling out (giving probability zero to) states that could not. In this way, the system's mechanism and state constitute information (about the system's previous state) in the classic sense of reduction of uncertainty or ignorance. More precisely, the system's mechanism and state generate 1 bit of information by distinguishing between things being one way (11 or 10, which remain indistinguishable to it) rather than another way (00 or 01, which also remain indistinguishable to it).

In general, the information generated when a system characterized by a certain mechanism in a particular state can be measured by the *relative entropy H* between the a posteriori and the a priori repertoires ("relative to" is indicated by $\|$), also known as *effective information* (*ei*):

$$ei(X_0 \to x_1) = H\left[p(X_0 \to x_1) \| \, p^{max} X_0\right]$$

Relative entropy, also known as Kullback-Leibler divergence, is a difference between probability distributions (Cover & Thomas, 2006): if the distributions are identical, relative entropy is zero; the more different they are, the higher the relative entropy. (Note that two different distributions over the same states have relative entropy >0 even if they have the same entropy.) Figuratively, the system's mechanism and state (01) generate information by sharpening the uniform distribution into a less uniform one—this is how much uncertainty is reduced. Clearly, the amount of effective information generated by a system is high if it has a large a priori repertoire and a small a posteriori repertoire, since a large number of initial states are ruled out. By contrast, the information generated is little if the system's repertoire is small or if many states could lead to the current outcome, since few states are ruled out. For instance, if noise dominates, so that any state could have led to the current one, no alternatives are ruled out, and no information is generated.

Since effective information is implicitly specified once a mechanism and state are specified, it can be considered to be an "intrinsic" property of a system. To calculate it explicitly, from an extrinsic perspective, one can perturb the system in all possible ways (i.e., try out all possible input states, corresponding to the maximum entropy distribution or a priori repertoire) to obtain a *forward* repertoire of output states given the system's mechanism, and finally calculate, using Bayes' rule, the a posteriori repertoire given the system's state (Balduzzi & Tononi, 2008).

Integration Second, we need to find out how much of the information generated by a system is integrated information—that is, how much information is generated by a single entity, as opposed to a collection of independent parts. The key idea here is to consider the parts of the system independently, ask how much information they generate by them-

selves, and compare it with the information generated by the system as a whole.

This can be done by resorting again to relative entropy to measure the difference between the probability distribution generated by the system as a whole (a posteriori of the system) with the probability distribution generated by the parts considered independently (the product of the a posteriori repertoire of the parts). Integrated information is indicated with the symbol ϕ (the vertical bar "I" stands for information, the circle "O" for integration (Tononi, 2004; Tononi & Sporns, 2003):

$$\phi(x_1) = H\left[p\left(X_0 \to x_1\right) \| \prod_{M^k \in \mu^{min}} p\left(M_0^k \to \mu_1^k\right) \right]$$

That is, the a posteriori repertoire for each part is specified by causal interactions internal to each part, considered as a system in its own right, while external inputs are treated as a source of extrinsic noise. The comparison is made with the particular decomposition of the system into parts that leaves the least information unaccounted for. This *minimum information partition* or $\mathbf{\wp}^{min}$ decomposes the system into its *minimal parts* (see Balduzzi and Tononi, 2008, for details).

To see how this system works, consider two of the million photodiodes in the digital camera (figure 84.2, *left*). By turning on or off depending on its input, each photodiode generates 1 bit of information, just as we saw before. Considered independently, then, 2 photodiodes generate 2 bits of information, and 1 million photodiodes generate 1 million bits of information. However, as shown in the figure, the product of the a posteriori distributions generated independently by the parts is identical to the a posteriori distribution for the system. Therefore, the relative entropy between the two distributions is zero: the system generates no integrated information [$\phi(x_1) = 0$] above and beyond what is generated by its parts.

Clearly, for integrated information to be high, a system must be connected in such a way that a lot of information is generated by causal interactions *among* its elements, rather than *within* its parts. Thus a system can generate integrated information only to the extent that it cannot be decomposed into informationally independent parts. A simple example of such a system is shown in figure 84.2 (*right*). In this case, the interaction between the minimal parts of the system generates information above and beyond what is accounted for by the parts by themselves, and $\phi(x_1) > 0$. In short, integrated information captures the information generated by causal interactions in the whole, over and above the information generated independently by the parts.

Complexes Finally, by measuring ϕ values for all subsets of elements within a system, we can determine which subsets form *complexes*. Specifically, a complex X is a set of elements that generate integrated information ($\phi > 0$) that is not fully contained in some larger set of higher ϕ (figure 84.3). A complex, then, can be properly considered to form a single entity having its own, intrinsic "point of view" (as opposed to being treated as a single entity from an outside, extrinsic point of view). Since integrated information is generated *within* a complex and not outside its boundaries, experience is necessarily private and related to a single point of view or perspective (Tononi, 2004; Tononi & Edelman, 1998). A given physical system, such as a brain, is likely to contain more than one complex, many small ones with low ϕ values, and perhaps a few large ones (Tononi, 2004; Tononi & Edelman, 1998). In fact, at any given time there may be a single *main complex* of comparatively much higher ϕ that underlies the dominant experience (a main complex is such that its subsets have strictly lower ϕ). As shown in figure 84.3, a main complex can be embedded into larger complexes of lower ϕ: a complex can be causally connected, through *ports-in* and *ports-out*, to elements that are not part of it. According to the IIT, such elements can influence indirectly the state of the main complex without contributing directly to the conscious experience it generates (Balduzzi & Tononi, 2008; Tononi, 2004).

Accounting for Neurobiological Observations Can this approach account, at least in principle, for some of the basic facts about consciousness that have emerged from decades of clinical and neurobiological observations? Measuring ϕ and finding complexes is not easy for realistic systems, but it can be done for simple networks that bear some structural resemblance to different parts of the brain (Balduzzi & Tononi, 2008; Tononi, 2004).

For example, by using computer simulations, it is possible to show that high ϕ requires networks that conjoin functional specialization (due to its specialized connectivity, each element has a unique functional role within the network) with functional integration (there are many pathways for interactions among the elements; figure 84.4*A*.). In very rough terms, this kind of architecture is characteristic of the mammalian corticothalamic system: different parts of the cerebral cortex are specialized for different functions, yet a vast network of connections allows these parts to interact profusely. And indeed, as much neurological evidence indicates (Posner & Plum, 2007; Schiff, chapter 78, this volume), the corticothalamic system is precisely the part of the brain that cannot be severely impaired without loss of consciousness.

Conversely, ϕ is low for systems that are made up of small, quasi-independent modules (figure 84.4*B*; Balduzzi & Tononi, 2008; Tononi, 2004). This may be the reason why the cerebellum, despite its large number of neurons, does not contribute much to consciousness: its synaptic organization is such that individual patches of cerebellar cortex tend

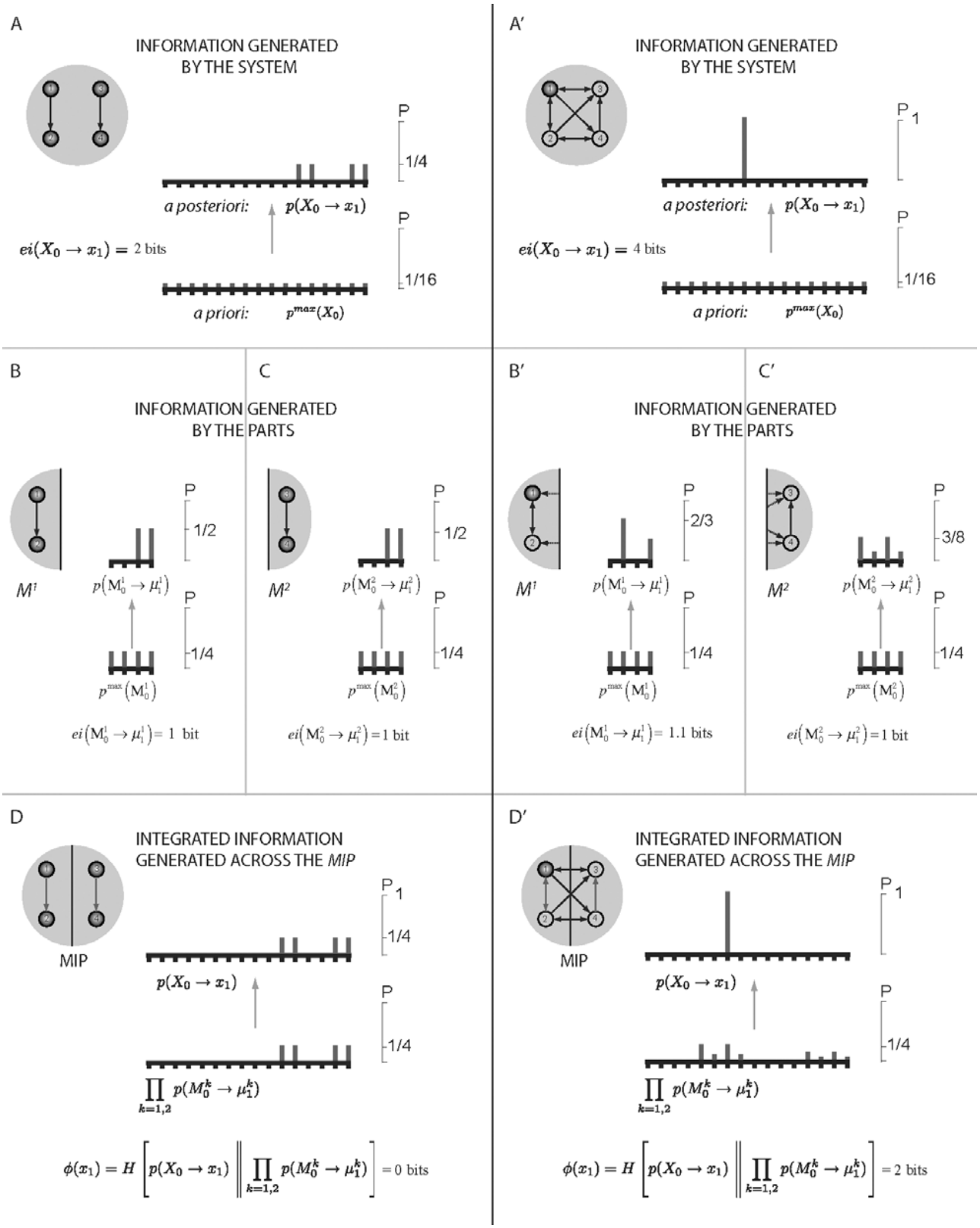

A
INFORMATION GENERATED BY THE SYSTEM
P
1/4
a posteriori: $p(X_0 \to x_1)$
P
$ei(X_0 \to x_1) = 2$ bits
1/16
a priori: $p^{max}(X_0)$
A′
INFORMATION GENERATED BY THE SYSTEM
P
1
a posteriori: $p(X_0 \to x_1)$
P
$ei(X_0 \to x_1) = 4$ bits
1/16
a priori: $p^{max}(X_0)$
B
INFORMATION GENERATED BY THE PARTS
P
1/2
M^1
$p(M_0^1 \to \mu_1^1)$
P
1/4
$p^{max}(M_0^1)$
$ei(M_0^1 \to \mu_1^1) = 1$ bit
C
P
1/2
M^2
$p(M_0^2 \to \mu_1^2)$
P
1/4
$p^{max}(M_0^2)$
$ei(M_0^2 \to \mu_1^2) = 1$ bit
B′
INFORMATION GENERATED BY THE PARTS
P
2/3
M^1
$p(M_0^1 \to \mu_1^1)$
P
1/4
$p^{max}(M_0^1)$
$ei(M_0^1 \to \mu_1^1) = 1.1$ bits
C′
P
3/8
M^2
$p(M_0^2 \to \mu_1^2)$
P
1/4
$p^{max}(M_0^2)$
$ei(M_0^2 \to \mu_1^2) = 1$ bit
D
INTEGRATED INFORMATION GENERATED ACROSS THE MIP
MIP
P
1
1/4
$p(X_0 \to x_1)$
P
1/4
$\prod_{k=1,2} p(M_0^k \to \mu_1^k)$
$\phi(x_1) = H\left[p(X_0 \to x_1) \,\middle\|\, \prod_{k=1,2} p(M_0^k \to \mu_1^k)\right] = 0$ bits
D′
INTEGRATED INFORMATION GENERATED ACROSS THE MIP
MIP
P
1
$p(X_0 \to x_1)$
P
1/4
$\prod_{k=1,2} p(M_0^k \to \mu_1^k)$
$\phi(x_1) = H\left[p(X_0 \to x_1) \,\middle\|\, \prod_{k=1,2} p(M_0^k \to \mu_1^k)\right] = 2$ bits

FIGURE 84.2 Integrated information. (*Left-hand side*) Two photodiodes in a digital camera. (*A*): the system as a whole generates 2 bits of effective information by specifying that n_1 and n_3 must have been on. (*B*,*C*): The information generated by the system as a whole is completely accounted for by the information generated independently by its parts. The minimum information partition (MIP) is the decomposition of a system into (minimal) parts that leaves the least information unaccounted for. (*D*) The a posteriori repertoire of the whole is identical to the combined a posteriori repertoires of the parts (the product of their respective probability distributions), so that relative entropy is zero. The system generates no information above and beyond the parts, so it cannot be considered a single entity. (*Right-hand side*) An integrated system. Elements in the system are on if they receive 2 or more spikes. The system is in state x_1 = 1000. (*A′*) The mechanism specifies a unique prior state that causes (leads to) state x_1, so the system generates 4 bits of effective information. All other initial states are ruled out, since they cause different outputs. (*B′C′*) Effective information generated by the two minimal parts, considered as systems in their own right. External inputs (dotted arrows) are treated as extrinsic noise and averaged over. (*D′*) The information generated by the whole (black arrows) over and above the parts (gray arrows). This is computed as the entropy of the a posteriori repertoire of the whole relative to the combined a posteriori repertoires of the parts: $\phi(x_1)$ = 2 bits. The system generates information above and beyond the parts, so it can be considered a single entity (a complex).

to be activated independently of one another, with little interaction between distant patches (Bower, 2002; Cohen & Yarom, 1998).

Computer simulations also show that units along multiple, segregated incoming or outgoing pathways are not incorporated within the repertoire of the main complex (figure 84.4*C*; Balduzzi & Tononi, 2008; Tononi, 2004). This may be the reason why neural activity in afferent pathways (perhaps as far as V1), though crucial for triggering this or that conscious experience, does not contribute directly to conscious experience; nor does activity in efferent pathways (perhaps starting with primary motor cortex), though it is crucial for reporting each different experience.

The addition of many parallel cycles also generally does not change the composition of the main complex, although

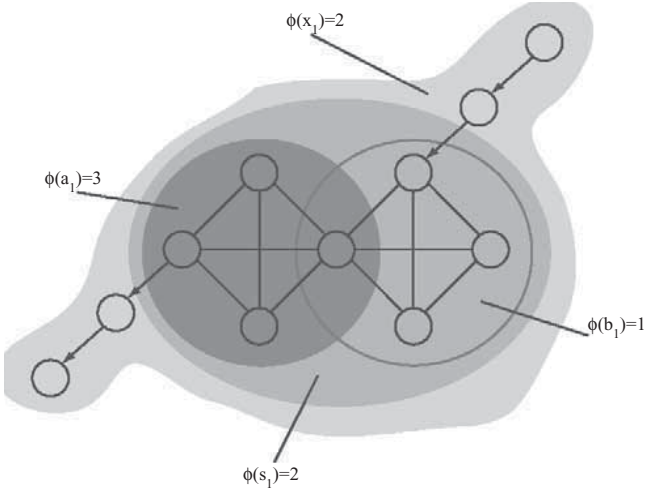


FIGURE 84.3 Complexes. Elements fire in response to an odd number of spikes; links without arrows are bidirectional. The system is decomposed into three of its complexes, shown in shades of gray. Observe that (1) complexes can overlap; (2) a complex can interact causally with elements not part of it; and (3) groups of elements with identical architectures generate different amounts of integrated information, depending on their ports-in and ports-out (compare subset *a*, the dark-gray filled-in circle, with the left side of subset *b*, the right-hand circle).

φ values can be altered (figure 84.4*D*). Instead, cortical and subcortical cycles or loops implement specialized subroutines that are capable of influencing the states of the main corticothalamic complex without joining it. Such informationally insulated cortical-subcortical loops could constitute the neural substrates for many unconscious processes that can affect and be affected by conscious experience (Baars, 1988; Tononi, 2004), such as those that enable object recognition, language parsing, or translating our vague intentions into the right words.

At this stage, it is hard to say precisely which cortical circuits may work as a large complex of high φ and which instead may remain informationally insulated. Does the dense mesial connectivity revealed by diffusion spectral imaging (Hagmann et al., 2008) constitute the "backbone" of a corticothalamic main complex? Do parallel loops through basal ganglia implement informationally insulated subroutines? Are primary sensory cortices organized like massive afferent pathways to a main complex higher up in the cortical hierarchy (Koch, 2004)? Is much of prefrontal cortex organized like a massive efferent pathway? Do certain cortical areas, such as those belonging to the dorsal visual stream, remain partly segregated from the main complex? Unfortunately, answering these questions and properly testing the predictions of the theory requires a much better understanding of cortical neuroanatomy than is presently available.

Other simulations show that the effects of cortical disconnections are readily captured in terms of integrated information (Tononi, 2004): a "callosal" cut produces, out of a large complex corresponding to the connected corticothalamic system, two separate complexes, in line with many studies of split-brain patients (Gazzaniga, 2005). However, because there is great redundancy between the two hemispheres, their φ value is not greatly reduced compared to when they form a single complex. Functional disconnections may also lead to a restriction of the neural substrate of consciousness, as is seen in neurological neglect phenomena, in psychiatric conversion and dissociative disorders, and possibly during dreaming and hypnosis. It is also likely that

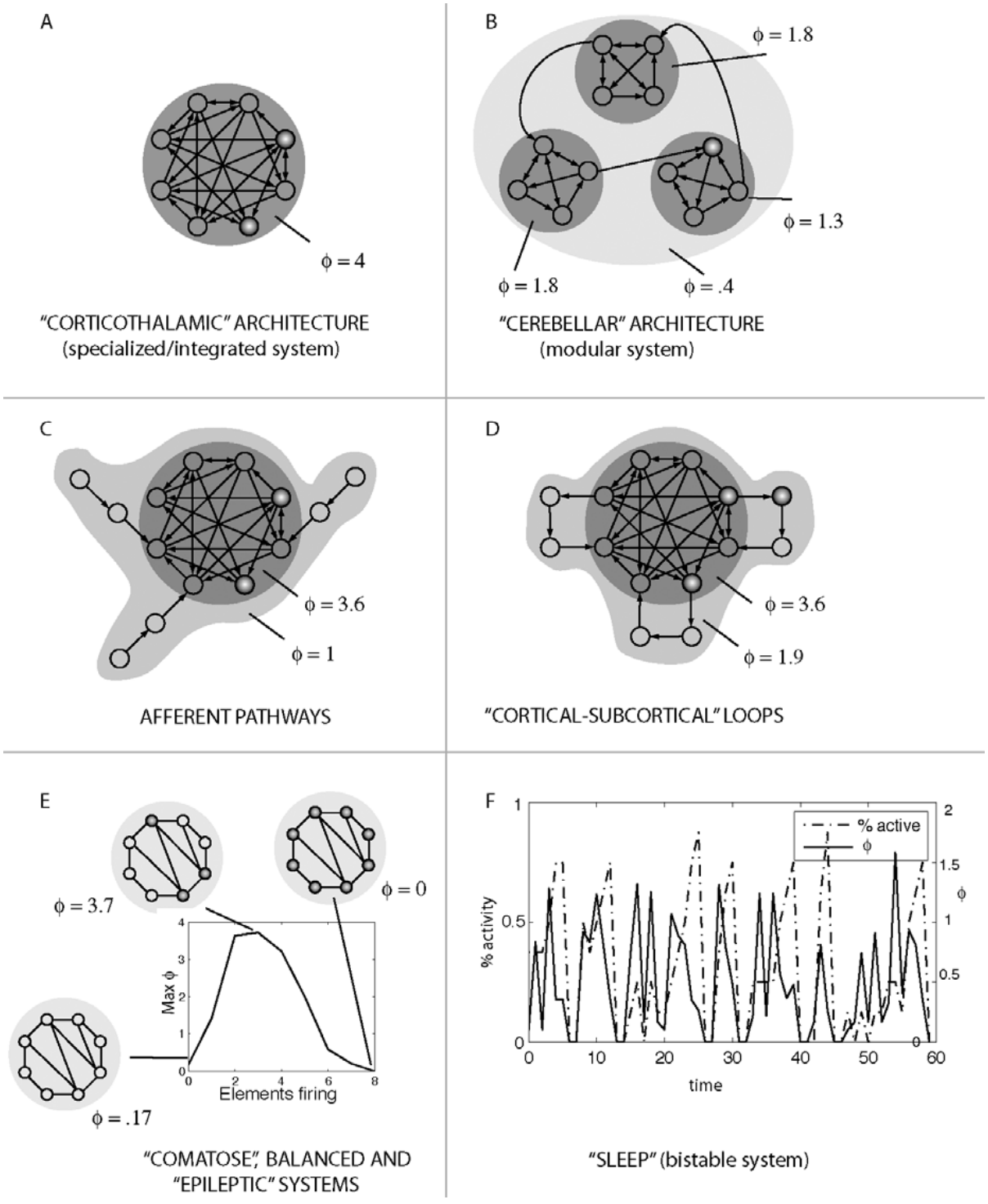


FIGURE 84.4 Relating integrated information to neuroanatomy and neurophysiology. Elements fire in response to two or more spikes (except elements targeted by a single connection, which copy their input); links without arrows are bidirectional. (*A*) Computing ϕ in simple models of neuroanatomy suggests that a functionally integrated and functionally specialized network—like the corticothalamic system—is well suited to generating high values of ϕ. (*B,C,D*) Architectures modeled on the cerebellum, afferent pathways, and cortical-subcortical loops give rise to complexes containing more elements, but with reduced ϕ compared to the main corticothalamic complex. (*E*) ϕ peaks in balanced states; if too many or too few elements are active, ϕ collapses. (*F*) In a bistable ("sleeping") system (same as in *E*), ϕ collapses when the number of firing elements (dotted line) is too high, remains low during the DOWN state, and only recovers at the onset of the next UP state.

certain attentional phenomena may correspond to changes in the composition of the main complex underlying consciousness (Koch & Tsuchiya, 2007; Koch, chapter 79, this volume). The attentional blink, where a fixed sensory input may at times make it to consciousness and at times not, may also be due to changes in functional connectivity: access to the main corticothalamic complex may be enabled or not based on dynamics intrinsic to the complex (Dehaene, Sergent, & Changeux, 2003). Similarly, binocular rivalry may be related, at least in part, to dynamic changes in the composition of the main corticothalamic complex caused by transient changes in functional connectivity (Lumer, 1998). Computer simulations confirm that functional disconnection can reduce the size of a complex and reduce its capacity to integrate information (Tononi, 2004). Although it is not easy to determine, at present, whether a particular group of neurons is excluded from the main complex because of hardwired anatomical constraints or is transiently disconnected as a result of functional changes, the set of elements underlying consciousness is not static, but forms a "dynamic complex" or "dynamic core" (Tononi & Edelman, 1998).

Computer simulations also indicate that the capacity to integrate information is reduced if neural activity is extremely high and near-synchronous, because of a dramatic decrease in the repertoire of discriminable states (figure 84.4*E*; Balduzzi & Tononi, 2008). This reduction in degrees of freedom could be the reason why consciousness is reduced or eliminated in absence seizure and other conditions during which neural activity is both high and synchronous.

The most common example of a marked change in the level of experience is the fading of consciousness that occurs during certain periods of sleep. Subjects awakened in deep NREM sleep, especially early in the night, often report that they were not aware of themselves or of anything else, though cortical and thalamic neurons remain active. Awakened at other times, mainly during REM sleep or during lighter periods of NREM sleep later in the night, they report dreams characterized by vivid images (Hobson & Pace-Schott, 2002; Hobson, Pace-Schott, & Stickgold, 2000). From the perspective of integrated information, a reduction of consciousness during early sleep would be consistent with the bistability of cortical circuits during deep NREM sleep. Because of changes in intrinsic and synaptic conductances triggered by neuromodulatory changes (e.g., low acetylcholine), cortical neurons cannot sustain firing for more than a few hundred milliseconds, and invariably enter a hyperpolarized down state. Shortly afterward, they inevitably return to a depolarized up state (Steriade, Timofeev, & Grenier, 2001). Indeed, computer simulations show that values of ϕ are low in systems with such bistable dynamics (figure 84.4*F*; Balduzzi & Tononi, 2008). Consistent with these observations, studies using TMS in conjunction with high-density EEG show that early NREM sleep is associated either with a breakdown of the effective connectivity among cortical areas, and thereby with a loss of integration (Massimini et al., 2007, 2005), or with a stereotypical global response suggestive of a loss of repertoire and thus of information (Massimini et al., 2007). Similar changes are seen in animal studies of anesthesia (Hudetz & Imas, 2007; Imas, Ropella, Ward, Wood, & Hudetz, 2005; Kroeger & Amzica, 2007).

Finally, consciousness not only requires a neural substrate with appropriate anatomical structure and appropriate physiological parameters: it also needs time (Bachmann, 2000). The theory predicts that the time requirements for the generation of conscious experience in the brain emerge directly from the time requirements for the buildup of an integrated repertoire among the elements of the corticothalamic main complex so that discriminations can be highly informative (Tononi, 2004; Balduzzi & Tononi, submitted). To give an obvious example, if one were to perturb half the elements of the main complex for less than a millisecond, no perturbations would produce any effect on the other half within this time window, and ϕ would be zero. After, say, 100 ms, however, there is enough time for differential effects to be manifested, and ϕ should grow.

The second problem: What determines the quality of consciousness?

Even if we were reasonably sure that a system is conscious, it is not immediately obvious what kind of consciousness it would have. For instance, our own consciousness comes in specific and seemingly irreducible qualities, exemplified by different modalities (vision, audition, pain, etc.), submodalities (visual color, motion, etc), and sub-submodalities (red, blue, etc.). What determines that colors look the way they do, and different from the way music sounds, or pain feels? Once again, neurological and neurophysiological evidence indicates that different qualities of consciousness must be contributed by different cortical areas. Thus damage to certain parts of the cerebral cortex impairs our ability to perceive visual motion, whereas damage to other parts selectively eliminates our ability to perceive colors (or to imagine, remember, and dream about them (van Zandvoort, Nijboer, & de Haan, 2007). There is obviously something about the organization or functioning of these cortical areas that makes them contribute specific qualities to conscious experience. Unless we accept that the kind of consciousness a system has is arbitrary, then, there must be some necessary and sufficient conditions that determine exactly what kind of experiences it can have. This is the second problem of consciousness.

THE SPECIFICITY OF CONSCIOUSNESS The intuitions that may help in addressing the second problem of consciousness build directly upon the approach taken to tackle the first

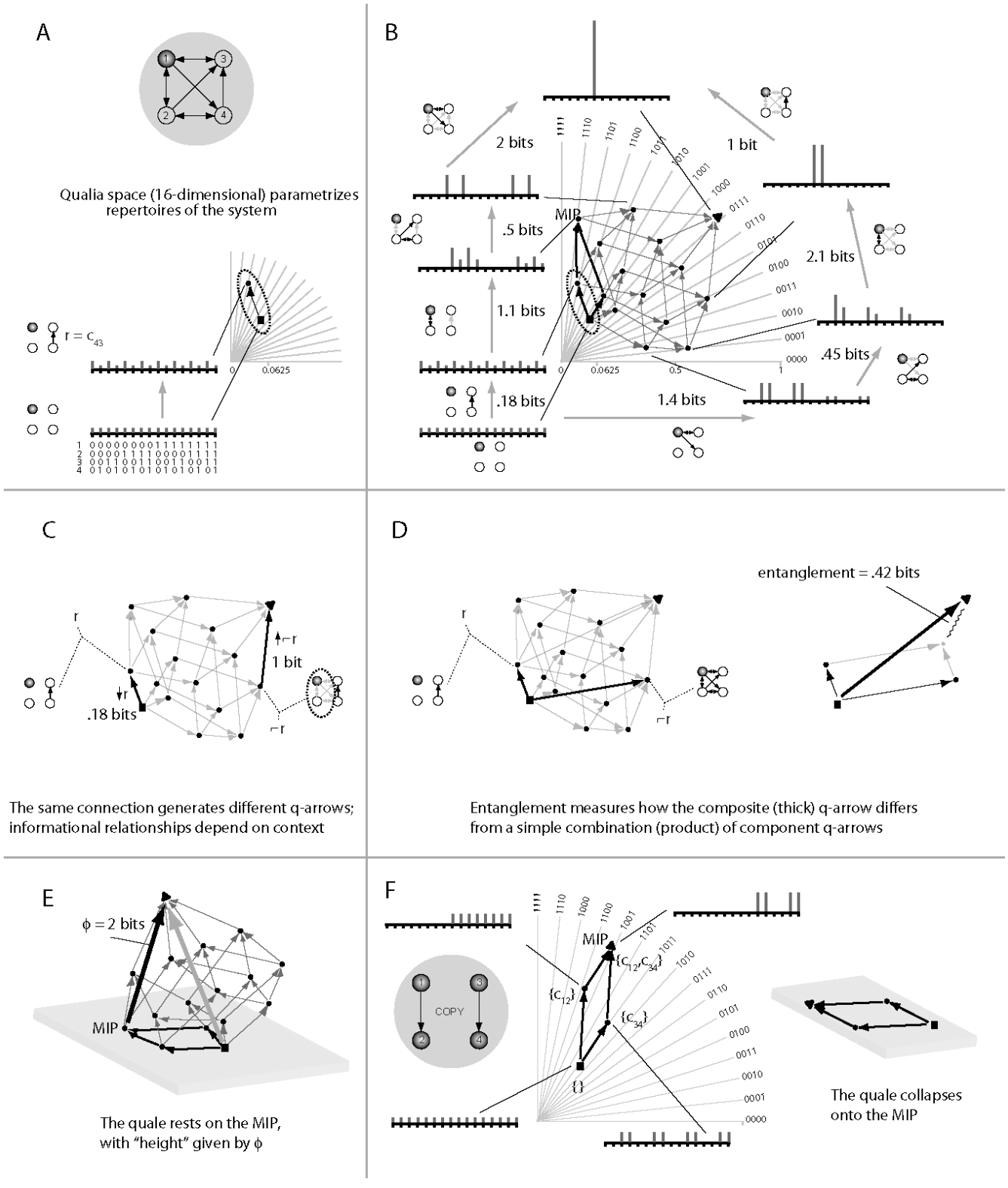

A
Qualia space (16-dimensional) parametrizes repertoires of the system
r = c43
0
0.0625
B
2 bits
1 bit
.5 bits
MIP
2.1 bits
1.1 bits
.45 bits
.18 bits
1.4 bits
0
0.0625
0.5
1
C
r
⌐r
1 bit
r
.18 bits
⌐r
The same connection generates different q-arrows; informational relationships depend on context
D
entanglement = .42 bits
r
⌐r
Entanglement measures how the composite (thick) q-arrow differs from a simple combination (product) of component q-arrows
E
ϕ = 2 bits
MIP
The quale rests on the MIP, with "height" given by ϕ
F
COPY
MIP
{c12, c34}
{c12}
{c34}
{}
The quale collapses onto the MIP

FIGURE 84.5 Qualia. (*A*): The system in the inset is the same as in figure 84.2*A′*. Qualia (*Q*) space for a system of 4 units is 16 dimensional (one axis per possible state; since axes are displayed flattened onto the page, and points and arrows cannot be properly drawn in two dimensions, their position and direction are for illustration purposes only). In state $x_1 = 1000$, the complex generates a quale or shape in *Q*-space, as follows. The maximum entropy distribution (the "bottom" of the quale, indicated by a black square) is a point assigning equal probability ($p = 1/16 = 0.0625$) to all 16 system states, close to the origin of the 16-dimensional space. Engaging a single connection *r* between elements 4 and 3 (c_{43}) specifies that, since element n_3 has not fired, the probability of element n_4 having fired the previous time step is reduced to $p = 0.25$ compared to its maximum ignorance value ($p = 0.5$), while the probability of n_4 not having fired is increased to $p = 0.75$. The a posteriori probability distribution of the 16 system states is modified accordingly. Thus the connection *r* "sharpens" the maximum entropy distribution into an a posteriori distribution, which is another point in *Q*-space. The *q-arrow* linking the two distributions geometrically realizes the *informational relationship* specified by the connection. The length (divergence) of the *q*-arrow expresses *how much* the connection specifies the distribution (the effective information it generates or relative entropy between the two distributions); the direction in *Q*-space expresses *the particular way* in which the connection specifies the distribution. (*B*) Engaging more connections further sharpens the a posteriori repertoire, specifying new points in *Q*-space and the corresponding *q*-arrows. The figure shows 16 out of the 399 points in the quale, generated by combinations of the four sets of connections. The insets around the quale are representative of the repertoires generated by two *q*-edges formed by *q*-arrows that engage the four sets of connections in two different orders (the two representative *q*-edges start at bottom left, one goes clockwise, the other counterclockwise; black connections represent those whose contribution is being evaluated; gray connections those whose contribution has already been considered and which provides the context on top of which the *q*-arrow generated by a black connection begins). Repertoires corresponding to certain points of the quale are shown alongside, as in previous figures. Effective information values (in bits) of the *q*-arrows in the two *q*-edges are shown alongside. Together, the *q*-edges enclose a shape, the quale, which completely specifies the quality of the experience. (*C*) The same connection considered in two contexts (black arrows). At the bottom of the quale (null context, corresponding to the maximum entropy distribution when no connections are engaged), the connection *r* generates a *q*-arrow (called down-set of *r*, or $\downarrow r$) corresponding to 0.18 bits of information pointing up-left in *Q*-space. Near the top of the quale (full context, corresponding to the a posteriori distribution specified by all other connections except for *r*, indicated as $\neg r$), *r* generates a *q*-arrow (called up-set of nonred, or $\uparrow \neg r$) corresponding to 1 bit of information pointing up-right in *Q*-space. (*D*) Entanglement. (*Left*) The *q*-arrow generated by the connection *r* and the *q*-arrow generated by the complementary connections $\neg r$ at the bottom of the quale (null context). (*Right*) The product of the two *q*-arrows (corresponding to independence between the informational relationships specified by the two sets of connections) would be a point corresponding to the vertex of the dotted parallelogram opposite to the bottom. However, "*r*" and "$\neg r$" jointly specify the a posteriori distribution corresponding to the top of the quale (black triangle). The distance between the probability distribution in *Q*-space specified jointly by two sets of connections and their product distribution (squiggly arrow) is the entanglement between the two corresponding *q*-arrows (how much the composite *q*-arrow specifies above and beyond its component *q*-arrows). (*E*) The *q*-edges converging on the minimum information partition of the system (MIP) form the natural base on which the complex rests, depicted as a "tent." The informational relationships *among* the parts are built on top of the informational relationships generated independently *within* the minimal parts. From this perspective the ϕ *q*-arrow (in black) is simply the tent pole holding the quale up above its base; the length (divergence) of the pole expresses the breathing room in the system. (*F*) The quale (not) generated by the double couple considered as a single system. Note that in this case the quale reduces to the MIP: the "tent" collapses onto its base, so there is no breathing room for informational relationships within the system. As shown in figure 84.2*A*, the system reduces to two independent parts, so it does not exist as a single entity. The quale generated by each part considered in isolation does exist, corresponding to an identical *q*-arrow for each couple.

problem. As we have seen, according to the IIT, the *quantity* of consciousness generated by a complex is determined by the amount of integrated information its mechanisms and state generate as a whole, above and beyond its parts. By extension, the IIT claims that the *quality* of consciousness generated by the complex is determined by the set of all the informational relationships generated by its mechanisms and state. That is, *how* integrated information is generated within a complex determines not only the amount of consciousness it has, but also what kind of consciousness.

Consider again the photodiode thought experiment. As we discussed before, when the photodiode reacts to light, it can only tell that things are one way rather than another way. On the other hand, when we see "light," we discriminate against many more states of affairs, and thus generate much more information. In fact, we argued that "light" means what it means and becomes conscious "light" *by virtue of* being discriminated not just against dark, but also against any color, any shape, any combination of colors and shapes, any frame of every possible movie, any sound, smell, thought, and so on.

What needs to be emphasized at this point is that discriminating "light" against all these alternatives implies not just picking one thing out of "everything else" (an undifferentiated bunch), but distinguishing at once, in a specific way, between each and every alternative. Consider a very simple example: a binary counter capable of discriminating among the four numbers: 00, 01, 10, 11. When the counter says binary "3," it is not just discriminating 11 from everything else as an undifferentiated bunch, otherwise it would not be a counter, but a 11 detector. To be a counter, the system must be able to tell 11 apart from 00, as well as from 10, as well as from 01 in different, specific ways. It does so, of course, by making choices through its mechanisms. For example, is this the first or the second digit? Is it a 0 or a 1? Each mechanism adds its specific contribution to the

discrimination they perform together. Similarly, when we see light, mechanisms in our brain are not just specifying "light" with respect to a bunch of undifferentiated alternatives. Rather, these mechanisms are specifying that light is what it is by virtue of being different, in this and that specific way, from every other alternative, from dark to any color to any shape, movie frame, sound or smell, and so on.

In short, generating a large amount of integrated information entails having a highly structured set of mechanisms that allow us to make many nested discriminations (choices) as a single entity. According to the IIT, these mechanisms working together generate integrated information by specifying a set of informational relationships that univocally determine the quality of experience.

Qualia Space: Giving Shape to Experience To see how this intuition can be given a mathematical formulation, let us consider again a complex X of n binary elements in a particular state x_1. Let us now suppose that each possible state of the system constitutes an axis or dimension of a *qualia space* (*Q*-space) having 2^n dimensions. Each axis is labeled with the probability p for that state, going from 0 to 1, so that a repertoire—that is, a probability distribution on the possible states of the complex—corresponds to a point in *Q*-space (figure 84.5).

Let us now examine how the connections among the elements of the complex specify probability distributions, that is, how a set of causal interactions specifies a set of informational relationships. First, consider the complex with all connections among its elements disengaged, thus discounting any causal interactions (figure 84.5*A*). In the absence of a mechanism, the state x_1 provides no information about the system's previous state: from the perspective of a system without causal interactions, all previous states are equally likely, corresponding to the maximum entropy or uniform distribution (the a priori repertoire). In *Q*-space, this probability distribution is a point projecting onto all axes at $p = 1/2^n$ (probabilities must sum to 1).

Next, consider engaging a single connection (figure 84.5*A*; the other connections are treated as extrinsic noise and averaged over). As with the photodiode, the mechanism implemented by that connection and the state the system is in rule out states that could not have caused x_1 and increases the a posteriori probability of states that could have caused x_1, yielding an a posteriori repertoire. In *Q*-space, the a posteriori repertoire specified by this connection corresponds to a point projecting onto higher p values on some axes and onto lower p values (or zero) on other axes. Thus the connection generates some information by shaping the uniform distribution into a more specific distribution, and thereby generates information (reduces uncertainty). More generally, we can say that the connection specifies an *informational relationship*, that is, a relationship between two probability distributions. This informational relationship can be represented as an arrow in *Q*-space (*q-arrow*) that goes from the point corresponding to the maximum entropy distribution ($p = \frac{1}{2}^n$) to the point corresponding to the a posteriori repertoire specified by that connection. The length (divergence) of the q-arrow expresses *how much* the connection specifies the distribution (the effective information it generates, i.e., the relative entropy between the two distributions); the direction in *Q*-space expresses *the particular way* in which the connection specifies the distribution, that is, a change in position in *Q*-space. Similarly, if one considers all other connections taken in isolation, each will specify another *q*-arrow of a certain length, pointing in a different direction.

Next, consider all possible combinations of connections (figure 84.5*B*). For instance, consider adding the contribution of the second connection to that of the first. Together, the first and second connections specify another a posteriori repertoire—another point in *Q*-space—and thereby generate more information than either connection alone as they shape the uniform distribution into a more specific distribution. To the tip of the *q*-arrow specified by the first connection, one can now add a *q*-arrow bent in the direction contributed by the second connection, forming an "edge" of two *q*-arrows in *Q*-space (the same final point is reached by adding the *q*-arrow due to the first connection on top of the *q*-arrow specified by the second one). Each combination of connection therefore specifies a *q-edge* made of concatenated *q*-arrows (component *q*-arrows). In general, the more connections one considers together, the more the a posteriori repertoire will take shape and differ from the uniform (a priori) distribution.

Finally, consider the joint contribution of all connections of the complex (figure 84.5*B*). As was discussed previously, all connections together specify the a posteriori repertoire of the whole. This is the point where all *q*-edges converge. Together, these *q*-edges in *Q*-space delimit a *quale*, that is, a *shape* in *Q*-space, a kind of 2^n-dimensional solid (technically, in more than 3 dimensions, the "body" of a polytope). The bottom of the quale is the maximum entropy distribution, its edges are *q*-edges made of concatenated *q*-arrows, and its top is the a posteriori repertoire of the complex as a whole. The shape of this solid (polytope) specifies all informational relationships that are generated within the complex by the interactions among its elements—also known as the *effective information matrix* (Tononi, 2004). Note that the same complex of elements, endowed with the same mechanism, will typically generate a different quale or shape in *Q*-space depending on the particular state it enters.

It is worth considering briefly two relevant properties of informational relationships or *q*-arrows. First, informational relationships are context-dependent (figure 84.5*C*) in the following sense. A *context* can be any point in *Q*-space corresponding to the a posteriori repertoire generated by a

particular subset of connections. It can be shown that the *q*-arrow generated by considering the effects of an additional connection (how it further changes the a posteriori repertoire) can change in both magnitude and direction depending on the context in which it is considered. In figure 84.5*C*, when considered in isolation (null context), the connection *r* between elements 4 and 3 generates a short *q*-arrow (0.18 bits) pointing in a certain direction. When considered in the full context provided by all other connections (not-*r* or ¬*r*), the same connection *r* generates a longer *q*-arrow (1 bit) pointing in a different direction.

Another important property of *q*-arrows is *entanglement* (Balduzzi & Tononi, submitted). Two Q-arrows are entangled if their underlying mechanisms, considered jointly, generate information above and beyond the information they generate separately (note the analogy with ϕ). Thus entanglement characterizes informational relationships (*q*-arrows) that are more than the sum of their component relationships (component *q*-arrows, figure 84.5*D*). Geometrically, entanglement "warps" the shape of the quale away from a simple hypercube (where *q*-arrows are orthogonal to each other). Entanglement has several relevant consequences (discussed in Balduzzi & Tononi, submitted), one of which being that it can identify *modes*. Modes are sets of *q*-arrows that are more densely entangled than surrounding *q*-arrows and can be considered as clusters of informational relationships constituting distinctive "subshapes" in *Q*-space.

Some Consequences of Viewing Qualia as Shapes What is the relevance of these constructs to understanding the quality of consciousness? It is not easy to become familiar with a complicated multidimensional space nearly impossible to draw, so it may be useful to resort to some metaphors (for a more detailed mathematical treatment, see Balduzzi & Tononi, submitted). Perhaps the most important notion emerging from this approach is that *an experience (a quale in the broad sense) is a shape in Q-space*. What gives each experience its particular shape are the informational relationships in *Q*-space (*q*-arrows between repertoires) generated by causal interactions among the elements of a complex. Only the informational relationships within a complex (those that give the quale its shape) contribute to experience. Conversely, the informational relationships that exist outside the main complex—for example, those involving sensory afferents—do not make it into the quale, and therefore do not contribute either to the quantity or to the quality of consciousness.

By the same token, different experiences are, literally, different shapes in *Q*-space. For example, when the same system is in a different state (firing pattern), it will typically generate a different shape or quale (even for the same value of ϕ). Moreover, experiences are similar if their shape is similar, and different to the extent that their shapes are different. Therefore, phenomenological similarities and differences can in principle be quantified as similarities and differences between shapes.

Note that a quale can only be specified by a mechanism and a particular state. On the one hand, it does not make sense to ask about the quale generated by a mechanism in isolation or by a state (firing pattern) in isolation. On the other hand, it does make sense to ask what kind of shapes or qualia the same system (mechanism) can generate when it is in different states. The set of all shapes generated by the same system in different states provides a geometrical depiction of all its possible experiences.

Another consequence is that two systems in the same state can generate two different experiences (i.e., two different shapes). As an extreme example, if a system were to copy one by one the state of the neurons in a human brain but had no internal connections of its own, it would generate no consciousness and no quale (Balduzzi & Tononi, 2008; Tononi, 2004).

Note also that informational relationships, and thus the shape of the quale, are specified both by the elements that are firing and by those that are not. This situation is natural considering that an element that does not fire will typically rule out some previous states of affairs (those that would have made it fire), and thereby it will contribute to specifying the a posteriori repertoire. Indeed, many silent elements can rule out, in combination, a vast number of previous states and thus be highly informative. From a neurophysiological point of view, such a corollary may lead to counterintuitive predictions. For example, take elements (neurons) within the main complex that happen to be silent when one is having a particular experience. If one were to temporarily disable them (e.g., make them *incapable* of firing), the prediction is that, though the system state (firing pattern) remains exactly the same, the quantity and quality of experience would change (Balduzzi & Tononi, 2008; Tononi, 2004).

It also follows that two systems with different architectures can generate the same experience (i.e., the same shape). For example, consider again the photodiode, whose mechanism determines that if the current in the sensor exceeds a threshold, the detector turns on. This simple causal interaction is all there is, and when the photodiode turns on, it merely specifies an a posteriori repertoire where states (00,01,10,11) have, respectively, probability (0,0,½,½). This corresponds in *Q*-space to a single *q*-arrow, one bit long, going from the a priori, maximum entropy repertoire (¼,¼,¼,¼) to (0,0,½,½). Now imagine the light sensor is substituted by a temperature sensor with the same threshold and dynamic range—we have a thermistor rather than a photodiode. While the physical device has changed, according to the IIT the experience, minimal as it is, has to be the same, since the informational relationship that is generated by the two devices is identical. Similarly, an AND gate when silent and

an OR gate when firing also generate the same shape in *Q*-space, and therefore must generate the same minimal experience (it can be shown that the two shapes are isomorphic, that is, have the same symmetries, Balduzzi and Tononi, submitted). In other words, different "physical" systems (possibly in different states) generate the same experience if the shape of the informational relationships they specify is the same. However, more complex networks of causal interactions are likely to create highly idiosyncratic shapes, so systems of high ϕ are unlikely to generate exactly identical experiences.

It is important to see what ϕ corresponds to in this representation (figure 84.5*E*). The minimum information partition (MIP) is just another point in *Q*-space: the one specified by the connections *within* the minimal parts only, leaving out the contribution of the connections *among* the parts. This point is the a posteriori repertoire corresponding to the product of the a posteriori repertoires of the parts taken independently. Then ϕ corresponds to an arrow linking this point to the top of the solid. In this view, the *q*-edges leading to the minimum information bipartition provide the natural "base" upon which the solid rests—the informational relationships generated *within* the parts upon which are built the informational relationships *among* the parts. The ϕ-arrow can then be thought of as the height of the solid—or rather, to employ another metaphor, as the highest pole holding up a tent. For example, if ϕ is zero (say a system decomposes into two independent complexes as in figure 84.5*F*), the tent corresponding to the system is flat—it has no shape—since the a posteriori repertoire of the system collapses onto its base (MIP). This configuration is precisely what it means when $\phi = 0$. Conversely, the higher the ϕ value of a complex, the higher the tent or solid, the more "breathing room" there is for the various informational relationships within the complex (the edges of the solid or the seams of the tent) to express themselves.

In summary, and not very rigorously, the generation of an experience can be thought of as the erection of a tent with a very complex structure: the edges are the tension lines generated by each connection in turn (the respective *q*-arrow or informational relationship). The tent literally takes shape when the connections are engaged and specify a posteriori repertoires. And when the system enters a different state, a different tent is erected.

Translating Phenomenology into Geometry The notions just sketched provide an initial framework for translating the seemingly ineffable qualitative properties of phenomenology into the language of mathematics, specifically, the language of informational relationships (*q*-arrows) in *Q*-space. Ideally, when sufficiently developed, such language should permit the geometric characterization of phenomenological properties generated by the human brain. In principle, it should also allow us to characterize the phenomenology of other systems. After all, in this framework the experience of a bat echolocating in a cave is just another shape in *Q*-space, and, at least in principle, shapes can be compared objectively.

At present, because of the combinatorial problems posed by deriving the shape of the quale produced by systems of just a few elements and because of the additional difficulties posed by representing such high-dimensional objects, the best one can hope for is to show that the language of *Q*-space can capture, in principle, some of the basic distinctions that can be made in our own phenomenology, as well as some key neuropsychological observations (Balduzzi and Tononi, submitted). A short list includes the following:

1. Experience is divided into modalities, like the classic senses of sight, hearing, touch, smell, and taste (and several others), as well as submodalities, like visual color and visual shape. What do these broad distinctions correspond to in *Q*-space? According to the IIT, modalities are sets of densely entangled *q*-arrows (*modes*) that form distinct *subshapes* in the quale; submodalities are subsets of even more densely entangled *q*-arrows (submodes) within a larger mode, thus forming distinct sub-subshapes. As a two-dimensional analog, imagine a given multimodal experience as the shape of the three-continent complex constituted by Europe, Asia, and Africa. The three continents are distinct subshapes, yet they are all part of the same landmass, just as modalities are parts of the same consciousness. Moreover, within each continent there are peninsulas (sub-subshapes), like Italy in Europe, just as there are submodalities within modalities.

2. Some experiences appear to be "elementary," in that they cannot be further decomposed. A typical example is what philosophers call a "quale" in the narrow sense, say, a pure color like red, or a pain, or an itch: it is difficult, if not impossible, to identify any further phenomenological structure within the experience of red. According to the IIT, such elementary experiences correspond to submodes that do not contain any more densely entangled sub-submodes (*elementary modes*).

3. Some experiences are homogeneous and others are composite: for example, a full-field experience of pure darkness, compared to that of a busy street. In *Q*-space, homogeneous experiences translate to a single homogeneous shape, and composite ones into a composite shape with many distinguishable subshapes (modes and submodes).

4. Some experiences are hierarchically organized. Take seeing a face: we see at once that as a whole it is somebody's face, but we also see that it has parts such as hair, eyes, nose, and mouth, and that those are made in turn of specifically oriented segments. The subjective experience is constructed from informational relationships (*q*-arrows) that are entangled (not reducible to a product of independent components)

across hierarchical levels. For example, informational relationships constituting "face" would be more densely tangled than unnatural combinations such as "eyelash + half of lip." The subshape of the quale corresponding to the experience of seeing a face is then an overlapping hierarchy of tangled *q*-arrows, embodying relationships within and across levels.

5. We recognize intuitively that the way we perceive taste, smell, and maybe color is organized phenomenologically in a "categorical" manner, quite different from, say, the "topographical" manner in which we perceive space in vision, audition, or touch. According to the IIT, these hard-to-articulate phenomenological differences correspond to different basic subshapes in *Q*-space, such as 2^n-dimensional *gridlike* structures and *pyramidlike* structures, which emerge naturally from the underlying neuroanatomy.

6. Some experiences are more alike than others. Blue is certainly different from red (and irreducible to red), but clearly it seems even more different from middle C on the oboe. In the IIT framework, in *Q*-space colors correspond to different subshapes of the same kind (say pyramids pointing in different directions) and sounds to very different subshapes (say tetrahedra). In principle, such subjective similarities and differences can be investigated by employing objective measures of *similarity between shapes* (e.g., considering the number and kinds of symmetries involved in specifying shapes that are generated in *Q*-space by different neuroanatomical circuits).

7. Experiences can be refined through learning and changes in connectivity. Say, for example, one learns to distinguish wine from water, then reds from whites, then different varietals. Presumably, underlying this phenomenological refinement is a neurobiological refinement: neurons that initially were connected indiscriminately to the same afferents become more specialized and split into subgroups with partially segregated afferents. This process has a straightforward equivalent in *Q*-space: the single *q*-arrow generated initially by those afferents *splits* into two or more *q*-arrows pointing in different directions, and the overall subshape of the quale is correspondingly refined.

8. Qualia in the narrow sense (experiential primitives) exist "at the top of experience" and not at its bottom. Consider the experience of seeing a pure color, such as red. The evidence suggests that the "neural correlate" (Crick & Koch, 2003) of color, including red, is probably a set of neurons and connections in the fusiform gyrus, maybe in area V8. (Ideally, neurons in this area are activated whenever a subject sees red and not otherwise, if stimulated trigger the experience of red, and if lesioned abolish the capacity to see red.) Certain achromatopsic subjects with dysfunctions in this general area seem to lack the feeling of what it is like to see color, its "coloredness," including the "redness" of red. They cannot experience, imagine, remember, and even dream of color, though they may talk about it, just as we could talk about echolocation, from a third-person perspective (van Zandvoort et al., 2007). Contrast such subjects, who are otherwise perfectly conscious, with vegetative patients, who are for all intents and purposes unconscious. Some of these patients may show behavioral and neurophysiological evidence for residual function in an isolated brain area (Posner & Plum, 2007; Schiff, chapter 78, this volume). Yet it seems highly unlikely that a vegetative patient with residual activity exclusively in V8 should enjoy the vivid perceptions of color just as we do, while being otherwise unconscious.

The IIT provides a straightforward account for this difference. To see how, consider again figure 84.5*C*: call "*r*" the connections targeting the "red" neurons in V8 that confer them their selectivity, and non-*r* ($\neg r$) all the other connections within the main corticothalamic complex. Adding *r* in isolation at the bottom of *Q*-space (null context), yields a small *q*-arrow (called the *down-set of red* or $\downarrow \neg r$) that points in a direction representing how *r* by itself shapes the maximum entropy distribution into an a posteriori repertoire. Schematically, this situation resembles that of a vegetative patient with V8 and its afferents intact but the rest of the corticothalamic system destroyed. The shape of the experience or quale reduces to this *q*-arrow, so its quantity is minimal (ϕ for this *q*-arrow is obviously low) and its quality minimally specified: as we have seen with the photodiode, *r* by itself cannot specify whether the experience is a color rather than something else, such as a shape, whether it is visual or not, sensory or not, and so on.

By contrast, subtract *r* from the set of all connection, so one is left with $\neg r$. This "lesion" collapses the *q*-arrow, called the *up-set of nonred* ($\uparrow \neg r$), which starts from the full context provided by all other connections $\neg r$ and reaches the top of the quale (more precisely, the lesion collapses all *q*-arrows generated by *r* starting from any *context*). This *q*-arrow will typically be much longer and point in a different direction than the *q*-arrow generated by *r* at the bottom of the quale. This result occurs because, the fuller the context, the more *r* can shape the a posteriori repertoire. Schematically, removing *r* from the top resembles the situation of an achromatopsic patient with a selective lesion of V8: the bulk of the experience or quale remains intact (ϕ remains high), but a noticeable feature of its shape collapses (the up-set of nonred). According to the IIT, the feature of the shape of the quale specified by "the up-set of nonred" captures the very quality or "redness" of red.

It is worth remarking that the last example also shows why specific qualities of consciousness, such as *the "redness" of red, while generated by a mechanism, cannot be reduced to a mechanism*. If an achromatopsic subject without the *r* connections lacks precisely the "redness" of red, whereas a vegetative patient with just the *r* connections is essentially unconscious, then

the redness of red cannot map directly to the mechanism implemented by the *r* connections. However, the redness of red can map nicely onto the informational relationships specified by *r*, as these change dramatically between the null context (vegetative patient) and the full context (achromatopsic subject).

Conclusion and outstanding issues

To recapitulate, the IIT claims that the quantity of consciousness is given by the integrated information (ϕ) generated by a complex of elements, and its quality by the shape in *Q*-space specified by all the informational relationships they generate. As we have tried to indicate, this theoretical framework can account, at least in principle, for several basic neurobiological and neuropsychological observations. Moreover, the same theoretical framework can be extended to begin translating phenomenology into the language of mathematics. At present, the very notion of a theoretical approach to consciousness may appear far-fetched, yet the nature of the problems posed by a science of consciousness seems to require a combination of experiments and theories: one could say that theories without experiments are lame, but experiments without theories are blind.

The IIT converges with other neurobiological frameworks (e.g., Crick & Koch, 2003; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Edelman, 1989) and cognitive theories (Baars, 1988) on several key facts: that our own consciousness is generated by distributed corticothalamic networks, that interactions among multiple cortical regions are important, and that there are many "unconscious" neural systems (or rather minimally conscious ones). Importantly, however, the IIT tries to account, in a coherent manner, for several key but puzzling facts about consciousness and the brain, such as the association of consciousness with the corticothalamic but not the cerebellar system or the fading of consciousness during certain stages of sleep or epilepsy despite continuing neural activity (as sketched in figure 84.4).

The IIT also predicts that consciousness depends exclusively on the ability of a system to generate integrated information. Whether or not the system is interacting with the environment on the sensory and motor side, it deploys language, capacity for reflection, attention, episodic memory, and a sense of space, of the body, and of the self. These are obviously important functions of complex brains and are crucial in shaping its connectivity. Nevertheless, contrary to some common intuitions but consistent with the overall neurological evidence, none of these functions seems absolutely necessary for the generation of consciousness "here and now" (Tononi & Laureys, in press).

Finally, the IIT says that the presence and extent of consciousness can be determined, in principle, also in cases in which we have no verbal report, such as infants or animals, or in neurological conditions such as minimally conscious states, akinetic mutism, psychomotor seizures, and sleepwalking. In practice, of course, measuring ϕ accurately in such systems will not be easy, but approximations and informed estimates are certainly conceivable. The theory also implies that consciousness is not an all-or-none property, but is graded: specifically, it increases in proportion to a system's repertoire of discriminable states. In fact, any physical system with some capacity for integrated information would have some degree of experience, irrespective of the constituents of which it is made and independent of its ability to report. In particular, this statement implies that suitably wired computers or robots can be conscious (Koch & Tononi, 2008). Whether these and other predictions turn out to be compatible with future clinical and experimental evidence, a coherent theoretical framework should at least help to systematize a number of neuropsychological and neurobiological results that might otherwise seem disparate.

To conclude, it is worth pointing out some outstanding issues that will need to be addressed in further developments of the theory. One of these is finding a principled way to determine the proper spatial and temporal scale to measure informational relationships and integrated information. What are the elements upon which probability distributions of states are to be evaluated? For example, are they synapses, neurons, or minicolumns? Similarly, what is the "clock" to use to identify system states? Does it run in milliseconds or hundreds of milliseconds? A working hypothesis is that the relevant spatial and temporal scales are those that jointly maximize ϕ (Tononi, 2004)—different systems will generate maximal amounts of integrated information at a particular spatiotemporal scale that is determined by their mechanism.

Another important issue has to do with the relationship between complexes and the outside world. The mechanisms of a complex generate integrated information and informational relationships from within. As shown by dreams, an adult brain does not need the outside world to generate experience. However, the mechanisms inside the complex are what they are, and so is the quality of the experience they generate, by virtue of a long evolutionary history, individual development, and learning. In fact, it appears that as a system incorporates statistical regularities from its environment and learns, its capacity for integrated information may grow (Tononi, Sporns, & Edelman, 1996). It will thus be important to see how the informational relationships (*q*-arrows) inside a complex reflect and react to informational relationships existing in the world. These and related issues, together with the intrinsic difficulty of characterizing *Q*-space for even the simplest of systems, will provide a challenging test bed for the IIT.

REFERENCES

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Bachmann, T. (2000). *Microgenetic approach to the conscious mind*. Amsterdam: John Benjamins.

Balduzzi, D., & Tononi, G. (2008). Integrated information in discrete dynamical systems: Motivation and theoretical framework. *PLoS Computat. Biol., 4*(6). doi:10.1371/journal.pcbi.1000091.

Balduzzi, D., & Tononi, G. (submitted). Qualia: The geometry of integrated information.

Bower, J. M. (2002). The organization of cerebellar cortical circuitry revisited: Implications for function. *Ann. NY Acad. Sci., 978*, 135–155.

Cohen, D., & Yarom, Y. (1998). Patches of synchronized activity in the cerebellar cortex evoked by mossy-fiber stimulation: Questioning the role of parallel fibers. *Proc. Natl. Acad. Sci. USA, 95*(25), 15032–15036.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nat. Neurosci., 6*(2), 119–126.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends Cogn. Sci., 10*(5), 204–211.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. USA, 100*(14), 8520–8525.

Edelman, G. M. (1989). *The remembered present: A biological theory of consciousness*. New York: Basic Books.

Gazzaniga, M. S. (2005). Forty-five years of split-brain research and still going strong. *Nat. Rev. Neurosci., 6*(8), 653–659.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol., 6*(7), e159.

Hobson, J. A., & Pace-Schott, E. F. (2002). The cognitive neuroscience of sleep: Neuronal systems, consciousness and learning. *Nat. Rev. Neurosci., 3*(9), 679–693.

Hobson, J. A., Pace-Schott, E. F., & Stickgold, R. (2000). Dreaming and the brain: Toward a cognitive neuroscience of conscious states. *Behav. Brain. Sci., 23*(6), 793–842; discussion, 904–1121.

Hudetz, A. G., & Imas, O. A. (2007). Burst activation of the cerebral cortex by flash stimuli during isoflurane anesthesia in rats. *Anesthesiology, 107*(6), 983–991.

Imas, O. A., Ropella, K. M., Ward, B. D., Wood, J. D., & Hudetz, A. G. (2005). Volatile anesthetics disrupt frontal-posterior recurrent information transfer at gamma frequencies in rat. *Neurosci. Lett., 387*(3), 145–150.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Denver, CO: Roberts.

Koch, C., & Tononi, G. (2008). Can machines be conscious? *Spectrum, IEEE, 45*(6), 55–59.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends Cogn. Sci., 11*(1), 16–22.

Kroeger, D., & Amzica, F. (2007). Hypersensitivity of the anesthesia-induced comatose brain. *J. Neurosci., 27*(39), 10597–10607.

Lumer, E. D. (1998). A neural model of binocular integration and rivalry based on the coordination of action-potential timing in primary visual cortex. *Cereb. Cortex, 8*(6), 553–561.

Massimini, M., Ferrarelli, F., Esser, S. K., Riedner, B. A., Huber, R., Murphy, M., et al. (2007). Triggering sleep slow waves by transcranial magnetic stimulation. *Proc. Natl. Acad. Sci. USA, 104*(20), 8496–8501.

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., & Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science, 309*(5744), 2228–2232.

Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. Cambridge, MA: MIT Press.

Posner, J. B., & Plum, F. (2007). *Plum and Posner's diagnosis of stupor and coma* (4th ed.). Oxford, UK: Oxford University Press.

Steriade, M., Timofeev, I., & Grenier, F. (2001). Natural waking and sleep states: A view from inside neocortical neurons. *J. Neurophysiol., 85*(5), 1969–1985.

Tononi, G. (2001). Information measures for conscious experience. *Arch. Ital. Biol., 139*(4), 367–371.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci., 5*(1), 42.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science, 282*(5395), 1846–1851.

Tononi, G., & Laureys, S. (in press). The neurology of consciousness: An overview. In S. Laureys & G. Tononi (Eds.), *The neurology of consciousness*. San Diego: Academic Press.

Tononi, G., & Sporns, O. (2003). Measuring information integration. *BMC Neurosci., 4*(1), 31.

Tononi, G., Sporns, O., & Edelman, G. M. (1996). A complexity measure for selective matching of signals by the brain. *Proc. Natl. Acad. Sci. USA, 93*(8), 3422–3427.

van Zandvoort, M. J., Nijboer, T. C., & de Haan, E. (2007). Developmental colour agnosia. *Cortex, 43*(6), 750–757.

XI

PERSPECTIVES

Chapter 85 BRUER 1221

86 BLUMSTEIN 1235

87 KOSSLYN, THOMPSON, AND GANIS 1241

88 GAZZANIGA, DORON, AND FUNK 1247

89 AMINOFF ET AL. 1255

85 Mapping Cognitive Neuroscience: Two-Dimensional Perspectives on Twenty Years of Cognitive Neuroscience Research

JOHN T. BRUER

Michael Gazzaniga and George A. Miller coined the name *cognitive neuroscience* in 1976, over martinis at the Rockefeller University Faculty Club. They chose the name to designate a new research program at the interface of systems neuroscience, computational neuroscience, and cognitive psychology. The goal of the research program would be to address the biological foundations of human cognition (Gazzaniga, 1984). In 1987, the James S. McDonnell Foundation, later joined by the Pew Trusts, made a commitment to grow and institutionalize this new field. One of the first initiatives McDonnell funded was the Summer Institute in Cognitive Neuroscience. Steve Kosslyn organized the first institute at Harvard University in 1988. Gazzaniga assumed the directorship of the Summer Institute the following year and has guided the institute since that time.

Gazzaniga initiated the practice of devoting every fifth Summer Institute to producing a volume that summarized the state of cognitive neuroscience at the time, presenting chapters that highlighted both progress made in the previous four years and outstanding research questions for the future. This volume, coming 20 years after the first Summer Institute, is the fourth such volume (Gazzaniga, 1995, 2000, 2004). These Summer Institute in Cognitive Neuroscience volumes (hereafter referred to as CN volumes) serve as significant resources for researchers and students in the field. Their impact comes as no surprise. Section editors and contributors were carefully picked from among the leading authorities in the field to present critical reviews of all areas of research, from cellular neuroscience to cognitive psychology, that were deemed relevant to the cognitive neuroscience enterprise.

As compilations of state of the science review articles, the CN volumes can be viewed not as snapshots, but rather as photo albums of the development of cognitive neuroscience. This perspective piece will capitalize on this feature of the volumes to initiate a historical look at this still relatively young field. Such a perspective is useful for established researchers and new students to assess progress the field has made, as well as to recall the origins of problems and questions in the field. Cognitive neuroscience also provides an interesting example for scholars of science to examine how an initially multidisciplinary research program coalesces into a new research field.

In this perspective piece, I attempt to initiate this historical discussion using articles published by contributors to the CN volumes as the starting point. As a first step, I will examine how publication patterns in cognitive neuroscience changed and how research topics changed between 1988 and 2007. Using bibliometric methods and data visualization techniques developed by information scientists, I will generate journal citation and topic word maps of a small portion of the cognitive neuroscience literature. Although the resulting maps are interesting and illustrative, keep in mind that they are first steps and generated from the published work of a relatively small and possibly unrepresentative sample of cognitive neuroscientists. Compared to larger studies of disciplines and maps of the entire scientific literature, the maps in this chapter can be characterized as "toy maps," in the same sense that early connectionist models were dubbed "toy networks." Like the early toy networks, these toy maps illustrate trends, questions, and possibilities that might be addressed in more extensive bibliometric and historical studies.

The author, publication, and topic word data set

Bibliometric studies are based on published documents and relationships that hold among them. Studies can be done at various levels of analysis (individual papers, authors, journals, institutions, nations, or scientific disciplines) and using various

JOHN T. BRUER James S. McDonnell Foundation, St. Louis, Missouri

relationships, among them: citation, cocitation, and coauthorship. This study will begin by compiling two author set. The first author set consists of the 150 authors who contributed chapters to CN 1995. The second author set consists of the 107 authors who have contributed to the current volume, CN 2008. These two author sets have 20 authors in common. For the 1995 contributors, I collected from the *Web of Science*® the bibliographic records for all articles these authors published in 1995. I will call this the *1995 data set.* To obtain a longer historical perspective, I also collected bibliographic records for all articles these authors published in 1988, the earliest year available on the *Web of Science*. I will call this the *1988 data set.* For the contributors to the current volume, I collected bibliographic records for all articles they published in 2007, the most recent complete year available on the *Web of Science* when the search was executed. This is the *2007 data se*t. The journals in which the authors published in each of the three data sets, plus data on journal cocitations, provide the basis for three journal citation maps (1988, 1995, and 2007). With these maps, one can visualize how publication patterns changed and see the citation flow among journals publishing cognitive neuroscientific research. Using programs developed by Loet Leydesdorff (Leydesdorff, 2004; Leydesdorff & Hellsten, 2005), I gathered title word cooccurrence data for articles in the three data sets to create topic word maps. With these maps, one can visualize how research topics in cognitive neuroscience and their interrelationships might have changed between 1988 and 2007. The data for the author sets are summarized in table 85.1.

The journal citation maps

A journal citation occurs when an article in journal A cites a previously published article in journal B. In citation maps, this relationship is depicted as A → B, which is read, "A cites B." Citations flow from journal A to journal B. If the journals cite each other, this is represented as A ↔ B. Journal citation maps have a long history in bibliometric studies of science. On a large scale, using hundreds or thousands of journals, they can be used to visualize interrelationships among major scientific disciplines. (For example, see www.eigenfactor.org.) On a smaller scale, they can be used to show how subdisciplines merge into new research fields. (Leydesdorff, 1994; McCain, 1998).

On the small scale employed in this study, one might hope to see how cognitive neuroscience emerged from its progenitor disciplines (systems neuroscience, cognitive psychology, neuropsychology) by noting changes in cocitation patterns among the progenitor discipline journals and possibly through the appearance of new cognitive neuroscience journals. These maps also allow us to visualize the citation flow among journals and to assess how results and ideas flowed among them. For an interdisciplinary field that is emerging from a multidisciplinary foundation, such as cognitive neuroscience, one might be able to see, for example, how ideas and results from neuroscience fed into psychology, from psychology into neuroscience, or both.

In order to ensure clear and readily interpretable maps, the journal citation maps presented here include only journals that published five or more articles in each of the data sets. There are around 30 journals in this category for each data set, and these journals contain on average 62% of the articles in each set (see table 85.1). These journals and the number of articles they published in each data set appear in table 85.2. These journals fall into four general categories: neuroscience (N), psychology (P), general (G), and clinical medicine (C).

The Science and *Social Science Citation Indexes* publish annual compilations of journal citation counts for major scientific journals, the *Journal Citation Reports* (JCR). For the 1988 journals, the citation data came from 1987 JCR, the year closest to 1988 for which I had access to hard copy volumes of the reports. For the 1995 and 2007 journals, data came from the on-line version of the JCR available through *Web of Science*: for 1995, the 1998 reports (the earliest year available on-line), and for 2007, the 2006 reports (the latest complete year at time of data collection).

For each of the data sets, the citation data is entered into an asymmetric matrix in which entries are the number of times row-journal A cites column-journal B. The matrix is asymmetrical because if journal A cites journal B *n* times, it is not generally the case that journal B cites journal A *n* times. Journal self-citations are omitted, so the matrix diagonal is empty. Asymmetric matrices are isomorphic to directed graphs, where in this case journals are nodes in the graph and directed edges represent the citation relation. It is these graph structures that information scientists call *maps*.

TABLE 85.1
Summary of 1988, 1995, and 2007 data sets

Year	Authors	Articles	Title Words	Journals	Journals with ≥5 Articles	Percentage of Articles in Journals with ≥5 Articles
1988	150	433	1531	144	26	57
1995	150	567	1870	159	33	64
2007	107	558	1713	146	27	66

TABLE 85.2

Journals containing five or more articles in each data set

1988		1995		2007	
Journal of Neuroscience (N)	21	*Investigative Ophthalmology and Visual Science* (C)	26	*Journal of Neuroscience* (N)	48
B Psychonom Soc (P)	15	*Journal of Neuroscience* (N)	25	*NeuroImage* (N)	29
Journal of Comparative Neurology (N)	15	*Journal of Neurophysiology* (N)	23	*Journal of Neurophysiology* (N)	22
Nature (G)	13	*Proceedings of the National Academy of Sciences of the U.S.A.* (G)	20	*Nature Neuroscience* (N)	21
Behavioural and Brain Sciences (P)	12	*Nature* (G)	19	*Neuropsychologia* (N)	20
Brain Research (N)	12	*Neuropsychologia* (N)	19	*Proceedings of the National Academy of Sciences of the U.S.A.* (G)	20
Proceedings of the National Academy of Sciences of the U.S.A. (G)	12	*Brain Cognition* (N)	15	*Cerebral Cortex* (N)	18
Science (G)	12	*Journal of Cognitive Neuroscience* (N)	14	*Neuron* (N)	18
Experimental Brain Research (N)	11	*Behavioural and Brain Sciences* (P)	13	*Sleep* (N)	15
Journal of Clinical and Experimental Neuropsychology (P)	11	*European Journal of Neuroscience* (N)	12	*Journal of Cognitive Neuroscience* (N)	14
Psychopharmacology (N)	9	*NeuroReport* (N)	12	*Biological Psychiatry* (C)	13
Trends in Neuroscience (N)	9	*Science* (G)	12	*Perception* (P)	13
Brain Cognition (N)	8	*Journal of Physiology (London)* (N)	11	*Science* (G)	13
Electroencephalograpy and Clinical Neurophysiology (C)	8	*Neuron* (N)	11	*Trends in Cognitive Science* (N)	12
Journal of Experimental Psychology: Learning, Memory, and Cognition (P)	8	*Journal of Comparative Neurology* (N)	10	*Nature* (G)	11
Journal of Neurophysiology (N)	8	*Trends in Neuroscience* (N)	10	*Brain Research* (N)	10
Psychophysiology (P)	8	*Cerebral Cortex* (N)	9	*Neuroscience Research* (N)	9
Cognition (P)	7	*Journal of Experimental Psychology: Human Perception and Perfomance* (P)	8	*Schizophrenia Bulletin* (C)	9
Cognitive Neuropsychology (P)	7	*Neuroscience* (N)	8	*Brain and Language* (P)	8
Epilepsia (N)	7	*Behavioural Neuroscience* (N)	7	*Journal of Vision* (C)	8
Neuropsychologia (N)	7	*Behavioural Brain Research* (N)	7	*Movement Disorders* (N)	7
Progress in Brain Research (N)	7	*Brain Research* (N)	7	*Neurology* (N)	6
Journal of Physiology (London) (N)	6	*Experimental Brain Research* (N)	7	*Psychophysiology* (P)	6
Behavioural Brain Research (N)	5	*Psychopharmacology* (N)	7	*Current Opinion in Neurobiology* (N)	5
Brain Research: Developmental Brain Research (N)	5	*Biological Psychiatry* (C)	6	*Experimental Brain Research* (N)	5
Perception (P)	5	*Current Opinion in Neurobiology* (N)	6	*Nature Reviews Neuroscience* (N)	5
		Molecular Brain Research (N)	6	*Der Nervenarzt* (C)	5
		Neurology (N)	6		
		Neuropsychology (N)	6		
		Personality and Individual Differences (P)	6		
		Journal of Experimental Psychology: Learning, Memory, and Cognition (P)	5		
		Psychological Science (P)	5		

Journals vary widely in the number of papers published in a year and thus vary in the citation opportunities they afford and in the number of citations they make. To normalize the citation data, the relative frequency with which journal A cites journal B (the number of times journal A cites journal B divided by the total number of citations journal A made in that year) is used as a measure of similarity or relevance between journals. The asymmetric relative frequency matrices are input to *Pajek*, a network analysis and visualization program (de Nooy, Mrvar, & Batagelj, 2005) that yields the directed graphs, or journal citation maps. To remove citation noise and clutter, a relative frequency threshold of greater than 0.03 is used. Edges that represent relative frequencies of 0.03 and less are removed from the maps. Nodes are placed in the map by using the Kamada-Kawai algorithm, which represents the network as a system of springs with relaxed lengths proportional to the edge length and iteratively repositions nodes to minimize overall energy of the spring system. The node size is scaled to the journal's "importance," which will be explained below. Isolated journals in the maps, that is, journals that do not cite or are not cited above the threshold by other journals, are shown in the top left of each map. Figure 85.1 presents the journal citation map for the 1988 data set.

Before analyzing entire maps, it is useful to focus first on portions of the maps, on subgraphs within the larger directed graphs. Is it possible to identify subsets of journals that mutually influence one another? Are there subsets of journals in which there is a citation flow from one journal in the subset to every other journal in the subset? If so, in a directed graph, these subsets of journals would form strong components of the graph. A strong component is the largest subset of nodes in the graph for which there is a directed path

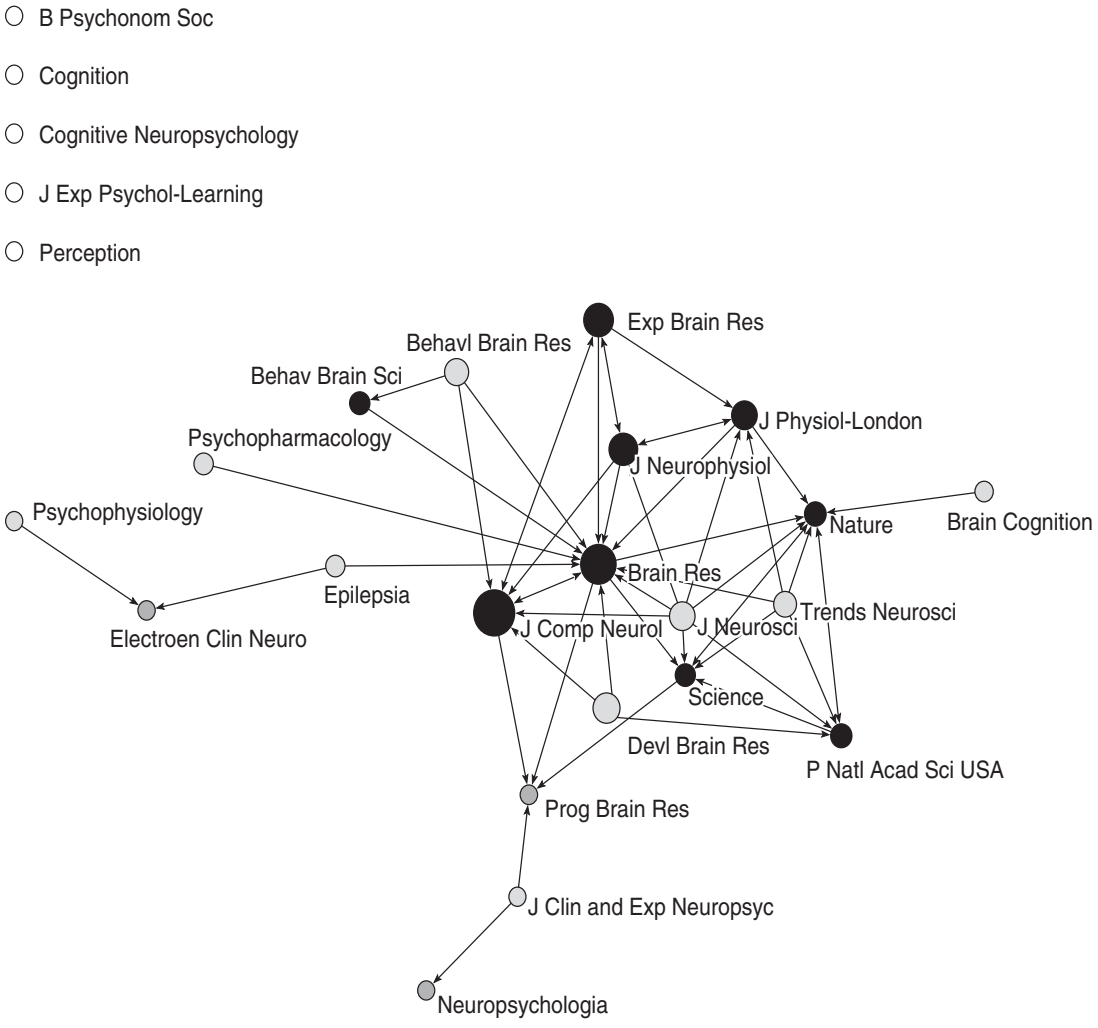


FIGURE 85.1 The 1988 journal citation map. Hub-authority journals are black nodes, authority journals are dark gray nodes, and hub journals are light gray nodes. Nodes are proportional to hub scores, authority scores, and hub + authority score for the black nodes.

(a path that follows the direction of the arrows) from any node in the subset to any other node in the subset. A strong component in the journal citation map is then the largest subset of journals for which there is a directed citation flow from each journal in the component to any other journal in the component. One can thus identify the strong components as cohesive sets of journals that mutually seek ideas, methods, and results from one another. The strong components in the journal citation maps represent cohesive subdisciplines of cognitive neuroscience. Figure 85.2 shows the strong components of the journal citation maps for 1998 (figure 85.2*A*), 1995 (figure 85.2*B*), and 2007 (figure 85.2*C*).

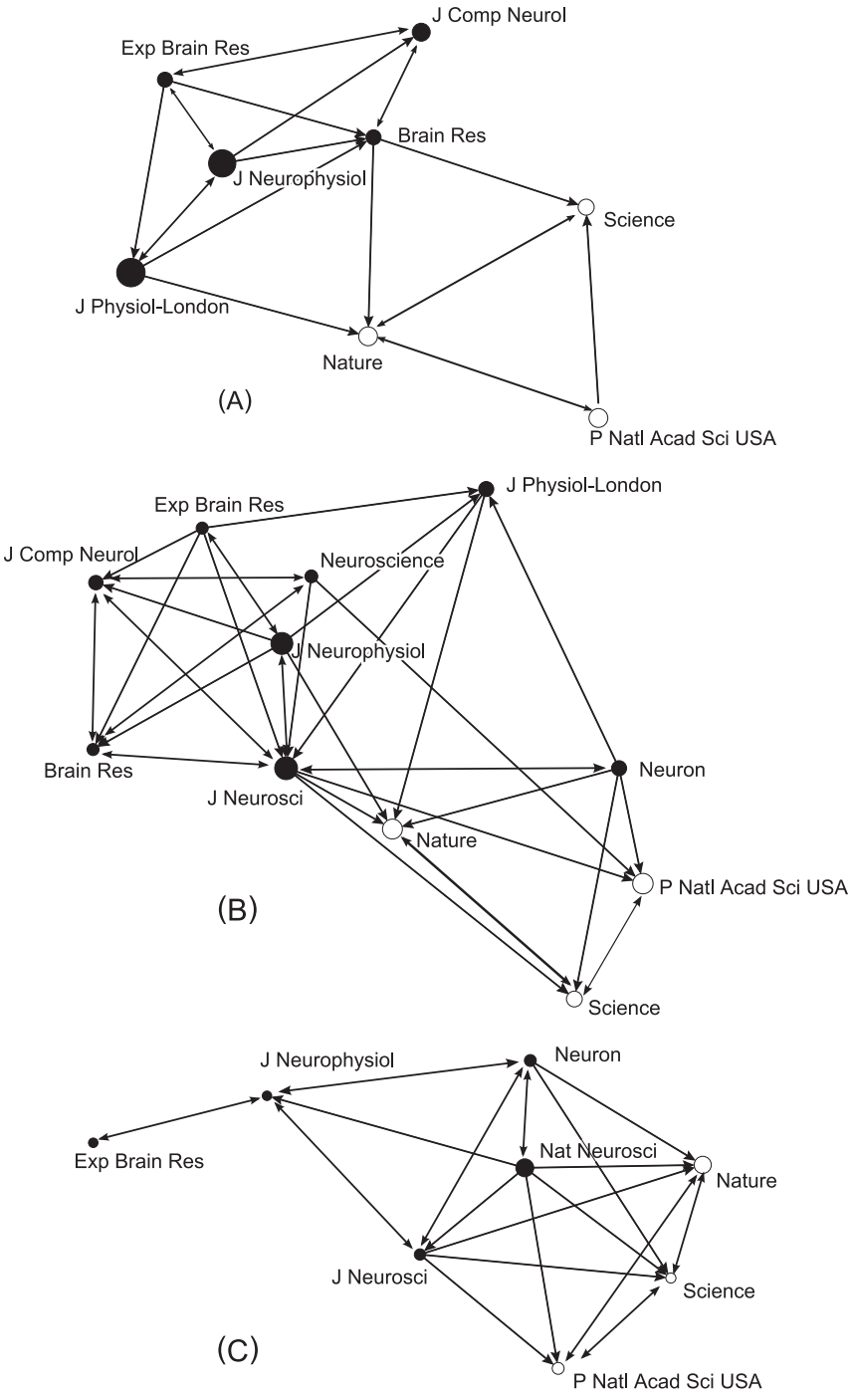


Figure 85.2 The neuroscience (black nodes) and general science (white nodes) strong components of the (*A*) 1988, (*B*) 1995, and (*C*) 2007 journal citation maps. Citations flow from neuroscience to general science strong component.

Each of the three maps contains the same two strong components. The larger component in each map, the neuroscience component (the black nodes in figure 85.2), contains journals that are categorized in table 85.2 as neuroscience journals. The number of journals in the neuroscience component varies from five (1988, 2007) to eight (1995). The *Journal of Neurophysiology* and *Experimental Brain Research* appear in the neuroscience strong component for all three data sets. The second, and smaller component, is a general science component (the white nodes in figure 85.2) which remains constant in the three maps and contains the prestigious, multidisciplinary journals *Nature, Proceedings of the National Academy of the U.S.A. (PNAS),* and *Science*. The percentage of articles in the article sets published by journals in the neuroscience component is 12% in 1988, 18% in 1995, and 21% in 2007. The percentage of articles published in the prestigious general science journals is between 8% and 9% each year. These two strong components, containing at most 5% of the journals for each data set, publish around 30% of the articles written by volume contributors in each of the years. No psychology or clinical journals appear in any of the strong components.

We can interpret the neuroscience component as representing a highly cohesive, mutually influential set of journals that captures one subdiscipline of cognitive neuroscience over the last 20 years. The existence of the general science component indicates that contributors to cognitive neuroscience are publishing articles in this highly selective group of journals and that the contributors to CN 1995 and 2007, by publishing in these journals, are part of the scientific mainstream.

Note also that the citation flow between the two strong components is the same for the three data sets. Journals in the neuroscience component cite journals in the general science component, and the converse never occurs. Articles in the general science component serve as sources for ideas, results, and methods in neuroscience. This is no surprise. The elite journals in the general science component are highly selective and publish across all areas of science. One would expect that neuroscience articles that meet publication criteria for the elite journals would be cited by core neuroscience journals, such as the *Journal of Neurophysiology* and the *Journal of Neuroscience*. Conversely, one would not expect disciplinary journals, such as the neuroscience journals, to be cited with high relative frequency in general science journals, which publish articles across the scientific spectrum.

Let us now turn to interpreting entire journal citation maps, as shown in figures 85.1, 85.3, and 85.4. First, one can discern the strong components, discussed above, at the center of each of the three maps. In the maps, there are also journals (five in 1998, four in 1995, and two in 2007) that are not connected to any other journals above the 0.03 threshold. In the 1988 map, all five isolated journals are psychology journals; in 1995, there is a two-journal component containing two of the major psychology journals. In 2007, there is a three-journal component that consists of two neuroscience journals and a clinical journal.

The large single components contain primarily neuroscience and general science journals. In 1988, three psychology journals appear on the periphery of the large component: *Behavioural and Brain Science, Journal of Clinical and Experimental Neuropsychology*, and *Psychophysiology*. In 1995, none of the five psychology journals listed in table 85.2 appear in the large component. In 2007, the three psychology journals—*Brain and Language, Perception*, and *Psychophysiology Brain and Language*—appear on the edge of the dominant neuroscience–general science component.

How might one identify "important" journals in these maps? Graph theorists and social network analysts have developed numerous methods for determining centrality, or prestige, of nodes in a network (Wasserman & Faust, 1994). Here, following a suggestion by Börner, Chen, and Boyack (2003), I will identify important journals in the maps by determining each journal's hub and authority scores. (Kleinberg, 1999) In analyses of links between pages on the World Wide Web, Kleinberg observed that some pages were pointed to by many hyperlinks and that these pages tended to contain primary or authoritative information on a topic. He called such pages *authorities*. There were other pages that sometimes contained little primary content but pointed to numerous pages that did. He called such pages *hubs*. Kleinberg developed a method to compute authority and hub scores for nodes in a directed graph. This method formalizes the intuition that a good authority is pointed to by other good hubs and a good hub points to many good authorities.

In the context of a journal citation map, journals with high authority scores are journals that are highly cited by other highly citing journals. Journals with high authority scores, then, would tend to serve as sources for ideas, methods, and results for the journals that cite them. A journal with a high hub score cites many other authorities and can be viewed as serving a synthesizing function by bringing together ideas, methods, and results from numerous authority journals.

Pajek includes a function that computes authority and hub scores for directed graphs, such as the journal citation maps. It also partitions the maps into four disjointed sets of nodes. It is these partitions that are shown in the maps in figures 85.1, 85.3, and 85.4. Some journals are neither authorities nor hubs (white nodes in the maps); some are both authorities and hubs (black nodes); some are authorities only (dark gray); and some are hubs only (light gray). Nodes representing the journals are scaled according to their authority and/or hub scores, the white nodes representing a zero score on both measures. The *Pajek* routine requires that one specify

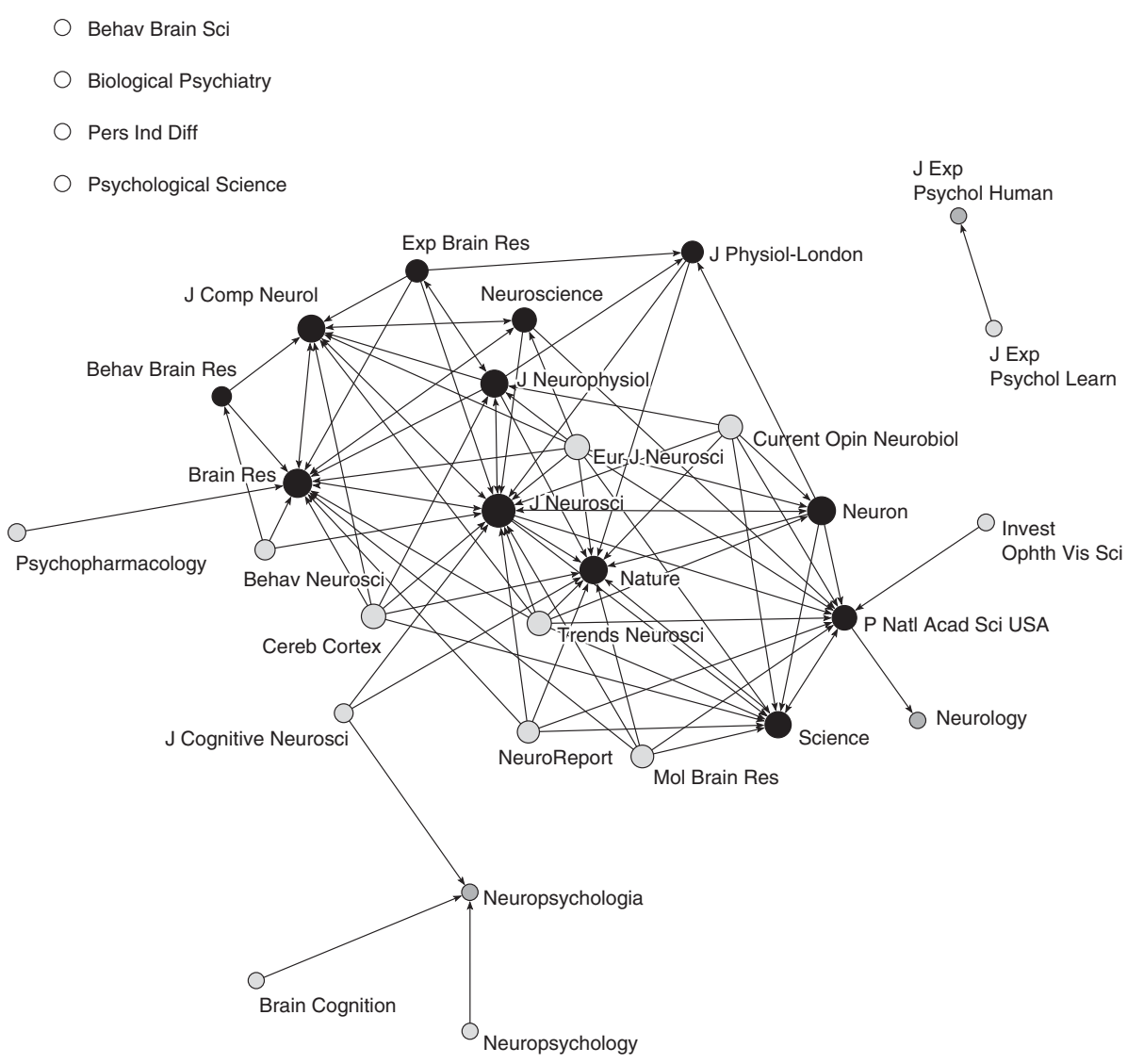


FIGURE 85.3 The 1995 journal citation map. Hub-authority journals are black nodes, authority journals are dark gray nodes, and hub journals are light gray nodes. Nodes are proportional to hub scores for light gray nodes, authority scores for dark gray nodes, and hub + authority score for the black nodes.

the number of authorities and hubs to be identified. I assumed that every cited journal is a potential authority and that every citing journal is a potential hub. Thus for 1988, the routine was requested to find 12 authorities and 18 hubs; for 1995, 15 authorities and 24 hubs; and for 2007, 14 authorities and 21 hubs.

The authority-hub journals with hub-plus-authority scores greater than 0.2 are shown in table 85.3. All the journals in table 85.3 are either neuroscience journals or general journals; that is, all journals that are in the two strong components are also both authorities and hubs. There are hub-authority journals that are not in one of the strong components, but their hub-plus-authority scores are quite low: *Behavioural and Brain Science* (score 0.10, a psychology journal) in the 1988 map, *Behavioral and Brain Research* (score, 0.11) in 1995, and *Brain Research* (score, 0.10) and *Neuropsychologia* (0.01) in 2007.

There are relatively few pure authority journals in the maps. The three pure authority journals in the 1988 map and the two in the 1995 are all neuroscience or clinical journals. In 2007, there are three pure authority journals: *Perception* and *Psychophysiology* are psychology journals, and *NeuroImage* is a neuroscience journal devoted to brain imaging studies and methods, which, as we will see in the topic maps, has become the mainstay of cognitive neuroscientific research. All these journals, with one exception, might be called degenerate pure authorities. They have authority scores less than 0.01; some have scores that are barely over zero. The exception is *NeuroImage* in the 2007 map, which has an authority score of 0.05.

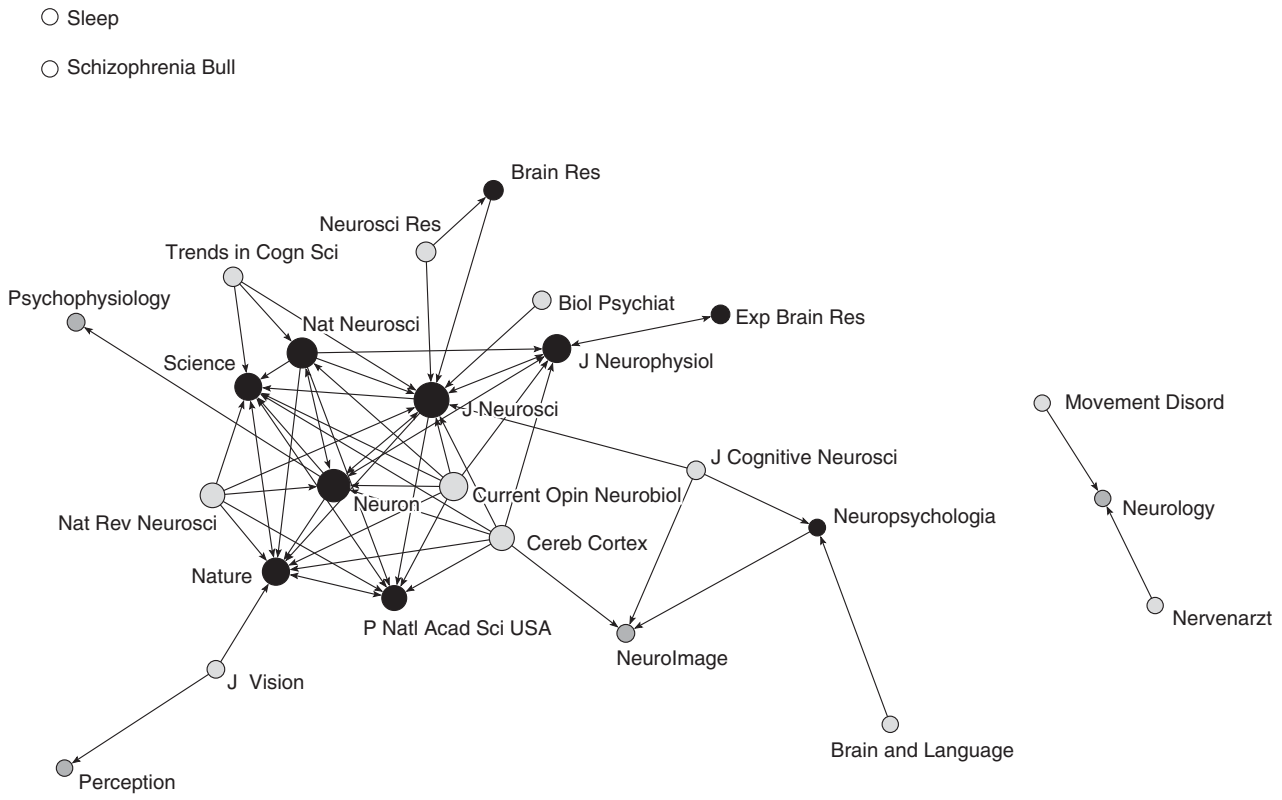


FIGURE 85.4 The 2007 journal citation map. Hub-authority journals are black nodes, authority journals are dark gray nodes, and hub journals are light gray nodes. Nodes are proportional to hub scores, authority scores, and hub + authority score for the nodes.

TABLE 85.3

Hub-authority journals with scores equaling the sum of their hub score and authority score

1988		1995		2007	
Journal of Comparative Neurology	1.19	*Journal of Neuroscience*	0.80	*Journal of Neuroscience*	0.89
Brain Research	0.82	*Brain Research*	0.52	*Neuron*	0.74
Experimental Brain Research	0.50	*Neuron*	0.50	*Nature Neuroscience*	0.60
Journal of Neurophysiology	0.45	*Nature*	0.47	*Journal of Neurophysiology*	0.48
Journal of Physiology—London	0.31	*Journal of Neurophysiology*	0.46	*Nature*	0.44
		Journal of Comparative Neurology	0.45	*Science*	0.43
		Science	0.42	*PNAS*	0.33
		PNAS	0.34		
		Neuroscience	0.31		
		Experimental Brain Research	0.23		
		Journal of Physiology—London	0.20		

The most common role for a journal in these citation maps is that of a pure hub. There are 9 pure hubs in the 1988 map, 13 in the 1995 map, and 11 in the 2007 map. Table 85.4 shows the pure hub journals with hub scores greater than 0.2. All are neuroscience journals and acquire their high hub score by dint of citing authoritative journals in both the neuroscience and general science strong components.

What is the relationship between psychology and neuroscience that is indicated by the publication patterns of the 1995 and 2008 CN contributors? If one looks at the number of psychology journals that publish more than five articles in our article sets, their number declines from eight psychology journals in 1988 to five psychology journals in 1995 to three psychology journals in 2007. The percentage of articles in each data set published in psychology

Table 85.4

Pure hub journals with hub scores greater than 0.2

1988		1995		2007	
Developmental Brain Research	0.37	*European Journal of Neuroscience*	0.39	*Current Opinion in Neurobiology*	0.50
Journal of Neuroscience	0.31	*Cerebral Cortex*	0.30	*Cerebral Cortex*	0.34
Behavioural Brain Research	0.25	*Molecular Brain Research*	0.28	*Nature Reviews Neuroscience*	0.32
Trends in Neurosciences	0.20	*NeuroReport*	0.28		
		Psychopharmacology	0.28		
		Trends in Neurosciences	0.28		
		Current Opinion in Neurobiology	0.27		

journals likewise decreases from 25.0% in 1988 to 10.3% in 1995 to 7.3% in 2007. On the basis of the overall exclusion of psychology journals from the main components of the journal citation maps and using authority and hub scores as measures of journal importance, one can conclude that the cognitive neuroscience literature, at least as published by contributors to the 1988 and 2008 CN volumes, is dominated by neuroscience and general science journals. There is no significant citation flow from neuroscience to psychology or conversely, at a relative frequency threshold of 0.03. If one goes below this threshold and includes all instances of a psychology journal citing a nonpsychology journal or conversely, for the 1995 and 2007 journal sets, in which complete data were available, there are 57 instances of a psychology journal citing a nonpsychology journal versus 20 instances of a nonpsychology journal citing a psychology journal. So for these authors and years, whatever citation flow there is appears to occur at a very low level and tends to be from psychology to neuroscience; that is, there is a greater tendency for psychology journals to look to neuroscience journals for ideas, results, and methods than conversely.

I mentioned in the introduction that in the late 1980s, computational neuroscience was also considered to be a contributing discipline to the development of cognitive neuroscience. In the 1558 articles in the three combined data sets, only four articles were published in dedicated computational neuroscience journals: two in *Computational Neuroscience* and two in *Neural Networks*. If computational neuroscience was a primary contributor to cognitive neuroscience, one can only assume that it is underrepresented among the authors and articles considered here. On the basis of these publication patterns, one can conclude cognitive neuroscience appears to be a variety of neuroscience and therefore that the field is appropriately named.

The journal citation maps also reveal something about the emergence of cognitive neuroscience as a field, apart from its relationship to progenitor disciplines. An important step in the development of any new field is a journal dedicated to work in the field. The *Journal of Cognitive Neuroscience* began publication in 1989 and appears in both the 1995 and 2007 journal citation maps. It published 2.5% of the entire 1995 article set (the eighth-ranked journal in number of articles published) and 2.5% of the 2007 article set (the tenth-ranked journal). In both maps, the *Journal of Cognitive Neuroscience* is a pure hub with moderate hub scores: 0.13 in 1995 and 0.07 in 2007. In this capacity, it appears to serve an interesting integrative function. In 1995, it appears to synthesize work from the *Journal of Neuroscience*, *Nature*, and *Neuropsychologia*. *Neuropsychologia*, according to its website, "publishes papers that explicitly address functional aspects of the brain." (www.elsevier.com) and describes itself as a journal in the behavioral and cognitive neurosciences. Functional aspects of the human brain as studied in neuropsychology rely heavily on cognitive psychological models of human behavior. As one can see in the 1995 map, *Neuropsychology* also cites *Neuropsychologia*. Thus in 1995, we can view the *Journal of Cognitive Neuroscience* as a journal integrating work in neuroscience, as found in the strong components of the citation maps, with neuropsychology and, through this connection to neuropsychology, to some extent with work in cognitive psychology.

In 2007, the *Journal of Cognitive Neuroscience* plays a similar integrating role. It once again synthesizes work from neuroscience and neuropsychology, as shown by its links to the *Journal of Neuroscience* and *Neuropsychologia*. However, it now also integrates ideas, methods, and results published in *NeuroImage*. *NeuroImage* began publication in 1993 and did not appear in the 1995 map. In 2007, however, it published 5.22% of the articles in that year's data set, second only to the *Journal of Neuroscience* (8.63% of the articles). Given that *Neuropsychologia* is the fifth-ranked journal in 2007, the *Journal of Cognitive Neuroscience* is a hub that cites and integrates work published in three of the most productive journals in the 2007 set.

Note also that *NeuroImage* is an authority journal with an authority score of 0.05. This journal is described as a journal that publishes imaging and modeling studies of structure-function relations in the brain (www.elsevier.com). The emerging prominence of this journal is indicative of the

central role that brain imaging technologies play in contemporary cognitive neuroscience. We will see more evidence of the centrality of imaging and recording studies in the next section on topic maps in cognitive neuroscience.

Topic maps

The stop-listed title words from articles contained in the three article sets can be used to generate topic maps of cognitive neuroscience for 1988, 1995, and 2007. (Freeware for compiling word co-occurrence matrices is referenced in Leytesdorff, 2004.) For each year, the analysis is limited to topic words that occur 11 or more times in the article titles for that year. There were 38 such words for the 1988 articles, 66 words for the 1995 articles, and 59 words for the 2007 articles. Co-occurrence matrices are symmetric, so the graphs, and therefore the maps, are undirected. In the matrix, each row is a vector of values giving the number of times the row title word occurs with the column title word. To normalize the data and to compute distances between topic words in the map, the cosine measure is used. This is the normalized inner product of the two vectors, which yields the cosine of the angle between the two vectors. The cosine measure varies from 0 (no similarity between the two topic-word vectors) to 1 (identity between the two topic word vectors). The maps below are drawn using a threshold of cosine of 0.2 or greater. Topic words are placed within the map using the Kamada-Kawai algorithm as explained above. In these maps, each node is scaled to the logarithm of the number of occurrences of the title word in the article set for that year. The smallest nodes in each map represent 11 occurrences. The largest node occurs in the 2007 map, for *cortex*, which occurs 76 times in article titles that year.

At a threshold of cosine ≥ 0.2, there are isolated nodes in the maps, words that despite their relatively high occurrence do not have vectors, or co-occurrence profiles, sufficiently similar to any other topic words to be linked to it in the map. In the 1988 map, there are 14 such isolated words; in the 1995 map, there are 21 such words; and in the 2007 map, there are 13 such words. These nodes have been removed from the maps.[1]

The three maps are shown in figure 85.5, 85.6, and 85.7. Before one looks for cohesive subsets of words, it is instructive to look at the overall structure of the maps. In each year, the most frequent topic words are *visual* and *cortex*, indicating the prominence of research on the visual system in neuroscience

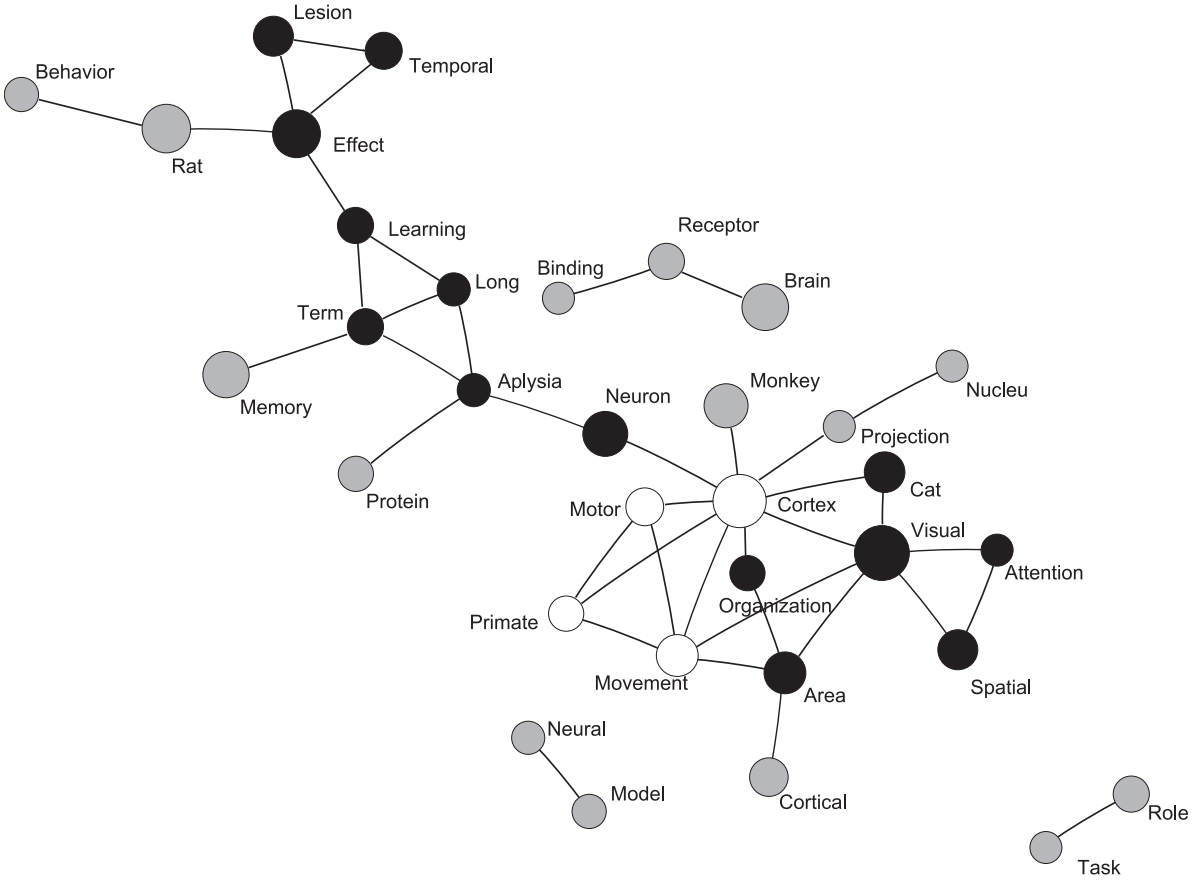


FIGURE 85.5 The 1988 cognitive neuroscience topic map. Black nodes are in the 2-core, and gray nodes are in the 1-core. Node size is proportional to the log of word occurrences.

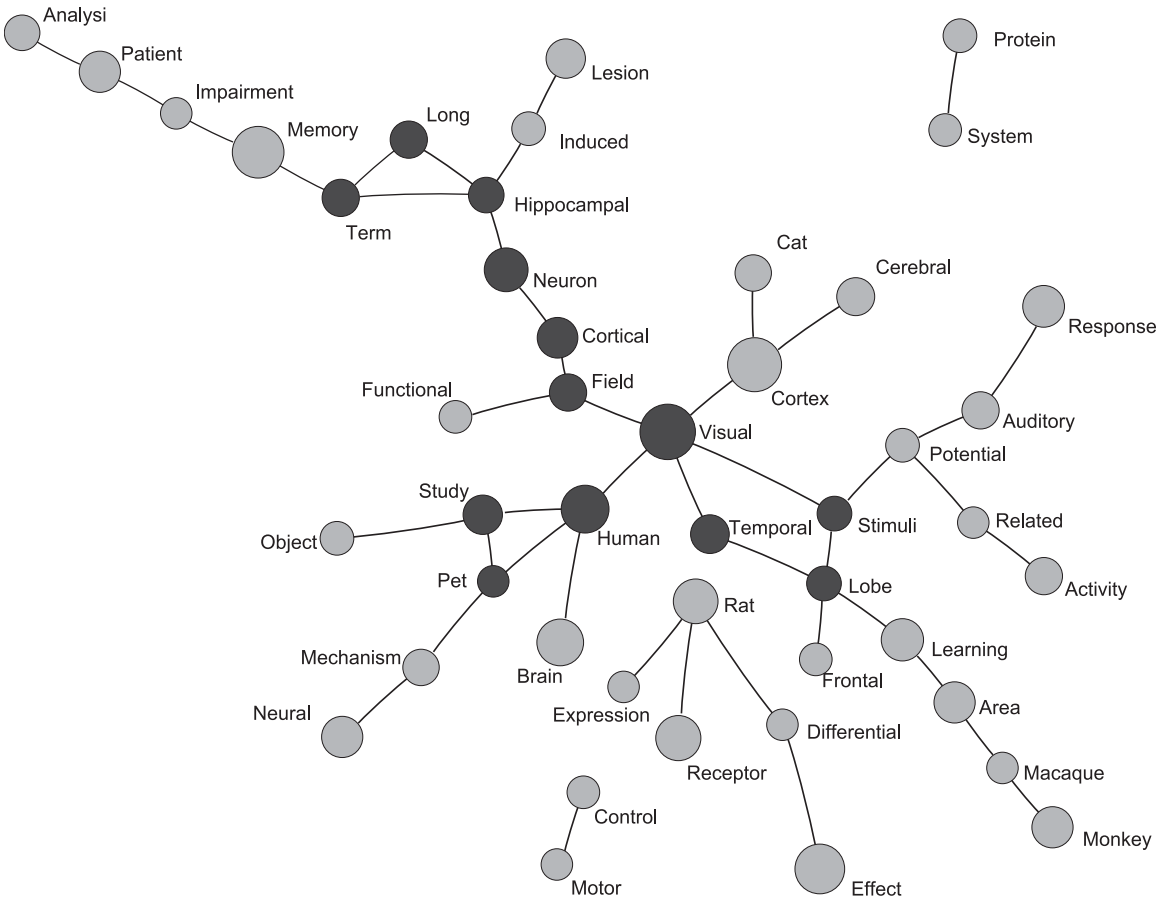


FIGURE 85.6 The 1995 cognitive neuroscience topic map. Black nodes are in the 2-core, and gray nodes are in the 1-core. Node size is proportional to the log of word occurrences.

and cognitive neuroscience. One can also see that the maps make intuitive sense, in that topic words in the same proximity correspond with research topics in cognitive neuroscience; for example, *visual-spatial-attention* in 1988, *hippocampal-long-term-memory* in 1995, and *transcranial-magnetic-stimulation* in 2007. The maps indicate the emergence of new methods, as with TMS, and new research areas, as shown in the *memory-emotion-amygdala* branch of the 2007 map.

One interesting change in the maps is in the topic words that refer to methods and experimental organisms. In the 1988 map, the only two words that refer to methods are *lesion* and *model*. In 1996, *lesion* is accompanied by *PET-study* and *response-potential*, the last two topic words referring to evoked response potential studies. By 2007, seven topic words, six of which refer to brain imaging and recording technologies, dominate the map. As for experimental organisms, in the 1988 map, five topic words referring to experimental organisms appear (*aplysia, cat, monkey, primate, rat*), none of which refer to humans. In 1995, four topic words designate nonhuman organisms (*rat, cat, macaque, monkey)*, but *human* and *patient* appear. In the 2007 map, only *human* and *patient* occur as referring to experimental organisms or, more accurately, to study participants. Over the 20-year period, method words in the maps increase, but the variety of experimental organisms dwindles to one: humans.

For each of the three years, the map contains one large connected component and several smaller disconnected components. In 1988, there are three disconnected components. The *model-neural* component might be interpreted as representing neural networks or connectionist models. *Binding-receptor-brain* might represent neurochemistry. *Task-role* is too vague to interpret. There are also three disconnected subnetworks in the 1995 map. The largest of these, *rat-expression-receptor-differential-effect*, might again represent neurochemistry or genetics. There are six disconnected subnetworks in the 2007 map that seem to refer to perception, neural correlates, cognitive control, motor control, object representations, and modeling.

Now let us look at the structure of the large connected components in each map. Cohesive subsets of topic words in these maps should delineate cohesive research topics that are prominent in cognitive neuroscience. One way to identify cohesive subsets of nodes in an undirected graph is to identify its *k*-cores. One can think of *k*-cores in the

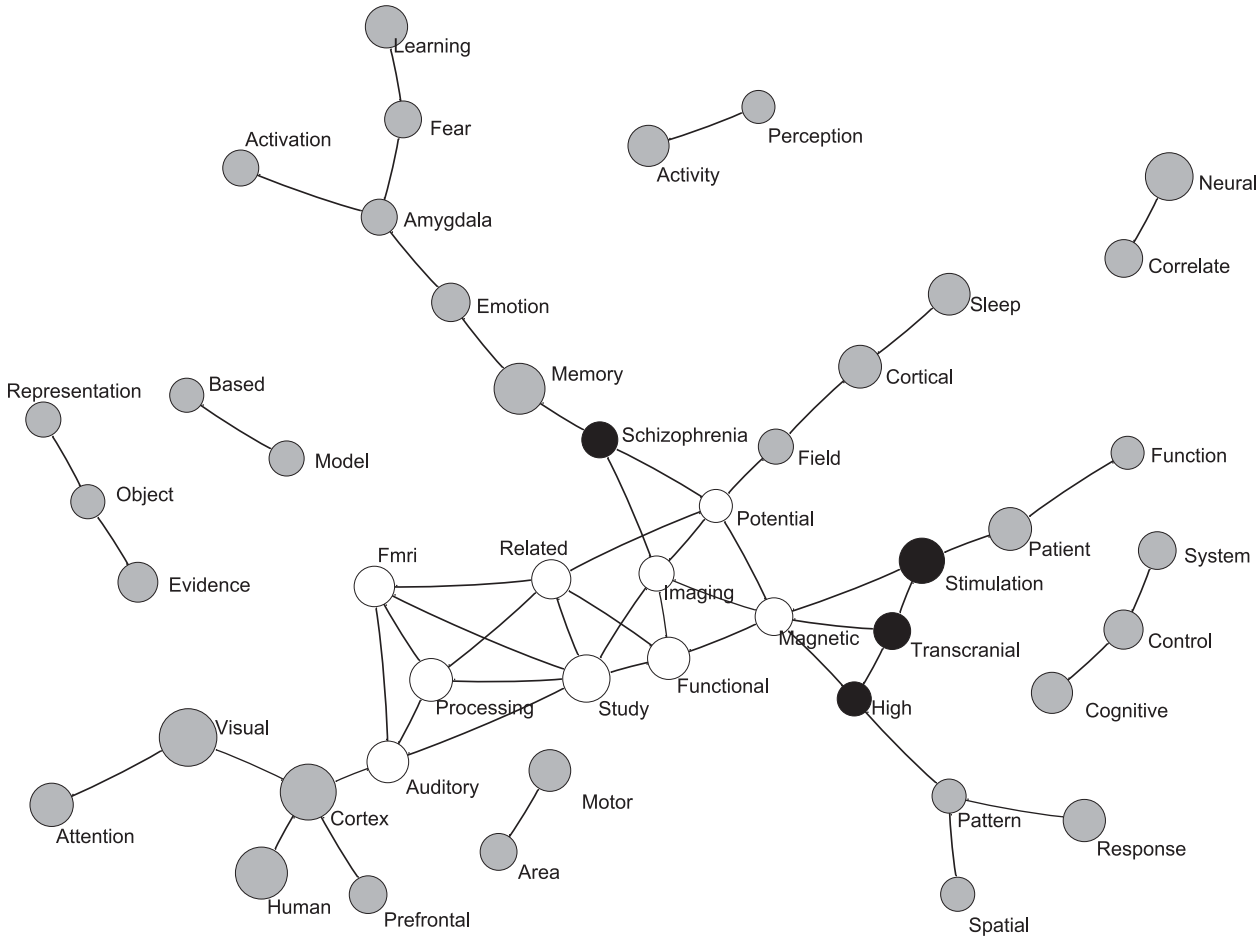


FIGURE 85.7 The 2007 cognitive neuroscience topic map. White nodes are in the 3-core, black nodes are in the 2-core, and gray nodes are in the 1-core. Node size is proportional to the log of word occurrences.

following way: Every topic word in the map is connected to at least one other topic word. Thus every node in the map is 1-connected or is a node in the map's 1-core. Some nodes of the 1-core are also connected to two other topic words; they are 2-connected. The largest set of 2-connected nodes is the map's 2-core. Of the set of all at least 2-connected nodes, some are 3-connected; the largest subset of these forms the 3-core. If we partition a map into its *k*-cores, we can envision nodes that only belong to a 1-core at the base of a mountain, where nodes belonging to higher-order cores form smaller layers of the mountain, until one arrives at the peak, the set of most densely interconnected nodes. (Think of the large connected component of the map as a wedding cake with three layers. The first layer is analogous to a 1-core. The plastic bride and groom stand atop the cake's 3-core.)

The nodes in the topic maps are shaded according to the *k*-core to which they belong. In these maps, the most densely connected topic words form a 3-core, shown as white nodes in the maps. Nodes in the 2-core are black, and nodes in the bottom 1-core are gray.

The peak of the 1988 map is a 3-core, which suggests that studies of motor control in primates was a highly cohesive research topic in that year. Note also that *cortex* is the most highly connected word in the map, having a co-occurrence profile that is sufficiently similar to those of nine other topic words to be linked to them in the map. In this map, the 3-core separates the 2-core. One part of the 2-core organizes around *visual*. *Visual* has the second highest number of connections in the map with six. The second part of the 2-core contains topic words relating to memory research.

In the 1995 map, a single 2-core, again organized around *visual*, serves as the backbone of the map. *Visual* once again has the highest number of connections: five. Extending out from vision are five branches, four of which are descriptive of research areas (memory, visual studies in cat, and temporal lobe studies). The fifth branch of the 2-core is a methodological branch indicating the emergence of human PET studies.

The *k*-core structure of the 2007 map is quite different. The peak in the map is a 3-core containing nine topic words, the majority of which refer to brain imaging and brain recording. *Magnetic* and *study* are each connected with six

other topic words, *imaging* and *related* with five. The 2-core is again separated into two noncontiguous parts, one representing the topic of transcranial magnetic stimulation (TMS) and one containing only the topic word *schizophrenia,* which in turn connects to 1-core words referring to memory, emotion, and other words descriptive of research on learning and affect. The two most common and highly connected topic words in previous maps, *visual* and *cortex*, are now part of the 1-core, with connections to only two and four other topic words, respectively. If we imagine the 3-core removed from the map, we are left with four independent branches of the map. One of the four branches is again a methods branch relating to TMS. The three remaining branches are research subject branches that contain words connoting affective processing, studies on sleep, visual attention and studies on prefrontal cortex.

Over the 20-year period, the topic maps change from being dominated by research areas with little mention of method to being dominated by topic words that refer to methods of brain imaging and recording. Overall, the number of method words appearing in the maps increases, and the number of experimental organisms decreases to include only human studies. Research topic words, such as *visual* and *cortex,* move from the peak of the maps to the base and decrease in their connectivity with other topic words. Brain imaging and recording come to occupy the highest ground in the maps and increase in their connectivity. This change in the topic maps is consistent with a change that we noted in the journal citation maps: *Neuorimage* emerged in the 2007 map as the second-ranked journal in number of publications and as a pure authority for three other journals in the map. Cognitive neuroscience appears to have changed from a collection of diverse research subjects to a field dominated, if not defined, by imaging technologies.

The change in the experimental organism from a variety of nonhuman animals to solely human also reflects a significant change in cognitive neuroscience research. The intent of most cognitive neuroscientists from the outset has been to understand human cognition, relying on animal models where needed and when they are the appropriate or only alternative. As Gazzaniga and Miller described the new field, its intent was to describe the biological foundations of human cognition. At the outset, in the early to mid-1980s, the methods for studying human cognition (with the exception of electroencephalography) were confined to behavioral studies using unimpaired (cognitive psychology) or impaired (neuropsychology) participant groups. Animal models provided the means for conducting invasive studies. Brain imaging technologies, particularly the development of PET and later fMRI and TMS, coupled with paradigms from cognitive psychology, allowed cognitive neuroscientists to map cognitive functions onto neural structures in normal human participants. It became possible to study, at one level at least, the biological foundations of human cognition in humans. From this perspective, the journal citation and topic maps reflect the development of the field over two decades into the discipline that Gazzaniga and Miller envisioned over martinis.

Conclusion

The CN volumes that I used in this study seem to reflect nicely the emergence of cognitive neuroscience over the past two decades. Although the samples used here are small, there is much more of scientific and historical interest that could be gleaned just from the contributions to these—now four—volumes. It would be presumptuous to state any strong conclusions based on the data and methods used here. So rather than offering conclusions, I will formulate two conjectures, which the data imply that others might test and debate. First, a positive conjecture: Noninvasive imaging and recording technologies have allowed cognitive neuroscience to develop into a science of the biological foundations of human cognition.

Second, a cautionary conjecture: Cognitive neuroscience not only has become coextensive with imaging studies, but also has become a variety of neuroscience, with psychology very much in the background. Might the full exploitation of advances in imaging technology require the constant infusion of better understandings of behavior, tasks, and task demands, along with better cognitive models?

In 1988, Michael Posner, Steve Petersen, Peter Fox, and Marcus Raichle (1988) articulated what I call the working hypothesis of cognitive neuroscience: "The human brain localizes mental operations of the kind posited by cognitive theories." (p. 1627) Imaging technology made this a viable and highly successful working hypothesis. This is consonant with my positive conjecture. In 1994, Posner and Raichle also stated, "The challenge for the future is to understand at a deeper level the actual mental operations assigned to the various areas of [brain] activation. Before this goal can be achieved, the experimental strategies used in PET studies must be refined so that more detailed components of the process can be isolated" (1994, p. 98).This statement would seem to suggest the importance of the continuous infusion of cognitive psychological ideas and results into cognitive neuroscience. This is consonant with my cautionary conjecture.

NOTE

1. The isolated topic words that were deleted from the maps and their number of occurrences are as follows: for 1988, amnesia (11), cell (17), cortical (18), development (19), evidence (12), human (23), pattern (17), patient (15), potential (12), processing (16), response (14), studies (14), study (13), and system (20) ; for 1995, attention (15), cell (22), cognitive (12), development (12), evidence (21), information (13), model (12), motion (13),

multiple (11), primate (12), priming (11), processing (15), regulation (11), representation (20), role (17), selective (13), spatial (26), specific (11), and task (17); for 2007, action (18), adult (11), behavioral (11), brain (40), effect (29), motion (16), movement (11), network (12), neuron (15), role (29), sensory (11), signal (11), and temporal (26).

REFERENCES

Börner, K., Chen, C., & Boyack, K. (2003). Visualizing knowledge domains. *Annu. Rev. Inform. Sci. Technol., 37*, 179–255.

de Nooy, W., Mrvar, A., Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. New York: Cambridge University Press.

Gazzaniga, M. S. (1984). *Handbook of cognitive neuroscience*. New York: Plenum Press.

Gazzaniga, M. S. (1995). *The cognitive neurosciences*. Cambridge, MA: MIT Press.

Gazzaniga, M. S. (2000). *The new cognitive neurosciences*. Cambridge, MA: MIT Press.

Gazzaniga, M. S. (2004). *The cognitive neurosciences* (3rd ed). Cambridge, MA: MIT Press.

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *J. ACM, 46*(5), 604–632.

Leydesdorff, L. (1994). The generation of aggregated journal-journal citation maps on the basis of the CD-ROM version of the Science Citation Index. *Scientometrics, 31*(1), 59–84.

Leydesdorff, L. (2004). The university-industry knowledge relationship: Analyzing patents and the science base of technologies. *J. Am. Soc. Inform. Sci. Technol., 55*(11), 991–1001.

Leydesdorff, L., & Hellsten, I. (2005). Metaphors and diaphors in science communication: Mapping the case of "stem-cell research." *Sci. Commun., 27*(1), 64–99.

McCain, K. W. (1998). Neural networks research in context: A longitudinal journal co-citation analysis of an emerging interdisciplinary field. *Scientometrics, 41*(3), 389–410.

Posner, M. I., Petersen, S. E., Fox, P. T., Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science, 240*(4859), 1627–1631.

Posner, M. I., & Raichle, M. (1994). *Images of mind*. New York: Scientific American Library.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University Press.

86 Reflections on the Cognitive Neuroscience of Language

SHEILA E. BLUMSTEIN

ABSTRACT This chapter provides some brief reflections on how the past 20 years of study of the cognitive neuroscience of language have changed the way in which we think about the nature of human language and the functional role of the two hemispheres in processing language. Two theoretical claims about the modularity of language are considered: (1) that language is modular in a narrow sense, that is, its various parts or components are functionally and neurally autonomous, and (2) that language is modular in a broad sense, that is, it is functionally and neurally separate from other cognitive functions. It is argued that the evidence to date challenges both claims of a theory of modularity. The functional properties of language, that is, speech, lexical processing, and syntactic processing, appear not to be focally represented in one area of the brain; rather, each recruits a broadly distributed neural network or processing stream. Moreover, certain areas of the brain that have been associated with language processing appear to be recruited across other cognitive domains, suggesting that while language may be functionally special, it draws on at least some neural mechanisms and computational properties shared across other cognitive domains. Finally, although it is generally assumed that the left hemisphere is dominant for language, functional neuroimaging studies often show activation in right hemisphere areas that are homologous to areas in the left hemisphere. These findings raise new questions about the potential role of the right hemisphere in language processing.

When we use language, it appears to be a unified whole in which all aspects—sounds, words, meaning, sentences, conversations—are integrated as a single piece. However, as we have known from the early historical record and the seminal work of Paul Broca and Carl Wernicke, language representation in the brain comprises a vast network in which damage to different areas of the brain have different linguistic consequences. Until about 20 years ago, the aphasias (speech and language impairments in adults as a consequence of organic brain pathology) served as the primary method of study and provided the foundation of our knowledge about brain-language relationships. Since then, a broader palette of methodologies has become available to study the cognitive neuroscience of language, such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), transcranial magnetic stimulation (TMS), and event-related potentials (ERP). These methods in conjunction with the lesion method have enriched and revolutionized the study of the neural basis of language and have provided both new insights and new challenges.

SHEILA BLUMSTEIN Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island

The goal of this chapter is to provide some brief reflections on how the past 20 years of study of the cognitive neuroscience of language have changed the way we think about the nature of human language itself and the functional role of the two hemispheres in processing language. In particular, we will focus on two theoretical claims about the modularity of language: that language is modular in a narrow sense, that is, its various parts or components are functionally and neurally autonomous, and that language is modular in a broad sense, that is, it is functionally and neurally separate from other cognitive functions. We will then briefly consider recent evidence that raises questions about the role of the right hemisphere in language processing.

The modularity of language: Components of the grammar

Owing to the rich history of study in linguistics and psycholinguistics, there is a general consensus about the general properties of language and the nature of language structure. The "pieces" that make up language have been operationally defined in terms of separate components including the sound structure of language (phonetics and phonology), lexical structure, morphological structure, and syntactic structure. Each of these components has separate functions and operates over different knowledge types (e.g., sounds, meanings). As a consequence, current models of language processing for speaking and understanding assume that language is a functionally modular system, that is, that each component is a separate module with a particular function (Pinker, 1994; Levelt, 1989).

Whether this means that for any given component of the grammar, there is a neural area dedicated to its processing is less clear. Past research from the aphasias has largely taken a modular approach, focusing on each component of the grammar separately and attempting to characterize deficits of patients in terms of impairments to a particular component of the grammar (Caplan, 1992, 1994; Grodzinsky,

Shapiro, & Swinney, 2000; Shelton & Caramazza, 1999). However, even the lesion data suggest that this story is not correct. For example, Broca's aphasics with damage including the inferior frontal gyrus (IFG) display deficits that cut across components of the grammar (cf. Grodzinsky & Amunts, 2006): They have phonetic and phonological impairments that affect particularly speech production but also speech perception (Blumstein, 2000); they have deficits in syntactic or sentence processing (Martin, Vuong, & Crowther, 2007; Grodzinsky, 1990), and they show impairments in lexical and semantic processing (Cappa & Perani, 2006; Milberg, Blumstein, Giovanello, & Misiurski, 2003). Wernicke's aphasics with damage including posterior portions of the superior temporal gyrus (STG) have auditory comprehension deficits that affect sound structure, lexical structure, and also syntactic structure (Blumstein, 2000; Yee, Blumstein, & Sedivy, 2008; Bates, Friederici, & Wulfeck, 1987; Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004; Piñango & Zurif, 2001). Additionally, these patients display impairments in spoken language production that cut across these components of the grammar (Blumstein, 2000; Yee et al., 2008; Baum, Blumstein, Naeser, & Palumbo, 1990; Goodglass, 1993; Faroqi-Shah & Thompson, 2003).

Because the lesions of aphasic patients tend to be very large, it is possible that multiple neural "modules" are indeed affected and contribute to the broad spectrum of language impairments of the patients. It is also possible that even within a particular neural area, there are functional subdivisions. Evidence from the neuroimaging literature can speak to these issues, allowing for a closer examination of the neural areas that are activated during various language tasks than that typically possible from lesion studies. There are some neuroimaging data that suggest functional subdivisions of the IFG across linguistic domains with Brodmann's area (BA) 45 involved in semantic processing and BA44 involved in phonological processing (Buckner, Raichle, & Petersen, 1995; Burton, 2001, 2009; Fiez, 1997; Poldrack et al., 1999), indicating that there might indeed be multiple "neural" modules relating to different functional properties of language. Nonetheless, there is currently much debate in the literature about whether different parts of the IFG reflect functional subdivisions of language or cut across these functional subdivisions and reflect different processes involved in cognitive control (Badre & Wagner, 2007; see also the section below entitled "The Modularity of Language: Language and Other Cognitive Functions"). Future studies will need to be done to determine how and in what ways different areas of the IFG contribute to language processing.

Whatever the role these different parts of the IFG may take, there is no question that many of the same areas (STG, middle temporal gyrus, supramarginal gyrus, and IFG) are involved in putatively different language functions. That is, these areas appear to be recruited in speech processing (Scott & Wise, 2004), lexical processing (Paulesu, Frith, & Frackowiak, 1993; Prabhakaran, Blumstein, Myers, & Hutchison, 2006), and syntactic processing (Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Kaan & Swaab, 2002). Thus the evidence to date does not support a strict modular view of language in which there is a "fixed neural architecture" (Fodor, 1983, p. 98) for each component of the grammar.

If this is the case, the challenge then is to determine the functional role of these different areas. Of course, it is possible that each component of the grammar has a dedicated neural architecture, but current neuroimaging methods cannot provide a sufficiently fine-grained picture of the neural structures associated with particular language functions. After all, activation of any particular voxel reflects responses of a large population of neurons that might themselves have distinctive and functionally distinct response properties. However, another possibility is that different neural areas have different computational properties, and the functional subdivisions do not cut across different information sources corresponding to the components of the grammar but rather reflect the nature of the computations that are done across these different information sources (cf. Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Poeppel, 2001; for further discussion, see the section below entitled "The Modularity of Language: Language and Other Cognitive Functions").

It is also clear from the neuroimaging literature that the functional properties of language, that is, speech, lexical, and syntactic processing, are not focally represented in one area of the brain; rather, each recruits a broadly distributed neural network or processing stream. An example from speech processing provides a window into the complexity of the neural systems underlying the processing of this one aspect or component of language.

The functional architecture of the speech perception system suggests a series of transformations of the acoustic input into a phonological representation. The neural instantiation of this system also appears to involve a hierarchical organization of the phonetic-processing stream in which information is transformed, conveyed, and ultimately acted upon with early auditory processing of the speech signal occurring in temporal areas and later stages of processing involving temporoparietal and frontotemporal systems. In particular, this phonetic processing stream includes temporal areas, including the STG, superior temporal sulcus (STS), and middle temporal gyrus (MTG); parietal areas, including the angular gyrus (AG) and supramarginal gyrus (SMG); and frontal areas, including the IFG (for detailed discussion, see Scott & Wise, 2004; Hickok & Poeppel, 2000, 2004). Nonetheless, the particular areas recruited vary as a function of the cognitive or task demands required (Poeppel, 1996). Mapping the acoustic-phonetic input to phonetic categories

recruits a different system from mapping this input to meaning or mapping that input to lexical form, and it recruits a different neural system for mapping acoustic-phonetic input to articulatory output. For example, one phonetic processing stream, the "what" stream, appears to be specialized for the recognition of auditory "objects" and recruits Heschl's gyri and the superior temporal lobes at early stages of processing, the left middle and anterior STG and STS of the left dominant hemisphere for the perception of speech sounds, and the left IFG for phonetic decisions (Scott & Wise, 2004; Hickok & Poeppel, 2004; Burton, Small, & Blumstein, 2000). The perception of phonetic category structure recruits the posterior STG, the SMG, and the IFG (Blumstein, Myers, & Rissman, 2005; Myers, 2007). The mapping of sound structure onto words involves a processing stream that involves temporal lobe structures, including the STG and the posterior portions of the STS and the MTG, as well as parietal lobe structures, including the AG and SMG (Paulesu et al., 1993; Prabhakaran et al., 2006). The mapping of auditory representations to articulatory representations recruits a processing stream that involves the STG, an auditory-motor interface area (the inferior parietal lobe including the posterior STG and the planum temporale) (Buchsbaum, Hickok, & Humphries, 2001; Hickok, Buchsbaum, Humphries, & Muftuler, 2003), and premotor cortex including the IFG and the SMA (Hickok & Poeppel, 2004).

These results suggest that the phonetic/phonological component of language and speech processing is not a modular system, involving solely dedicated neural machinery, but rather is a distributed neural system that requires the integration of multiple functional systems. Although phonetic processing itself might be functionally modular, it appears not to be neurally modular.

The modularity of language: language and other cognitive functions

There is ample evidence to show that there is neural specialization for different cognitive processes and that for most right-handers, language appears to be localized in the left hemisphere. Within the left hemisphere, language areas are located in areas surrounding the sylvian fissure, including Broca's area (the pars opercularis (BA44) and triangularis (BA45), Wernicke's area (the posterior superior and middle temporal gyrus), the angular gyrus (BA39), and the supramarginal gyrus (BA40). Lesions in any of these areas will cause an aphasia, and the locus of the lesion will result in a particular profile of language abilities and disabilities called a *symptom complex* (Goodglass, 1993).

It has also been shown that the neural systems underlying language can be dissociated from other cognitive functions and abilities. For example, patients with damage to different parts of the medial temporal lobe have memory deficits for recent events (anterograde amnesia) or for recalling the past (retrograde amnesia). Yet they show normal knowledge of language structure. Likewise, Alzheimer's disease patients who have neural degeneration in areas of the temporal and parietal lobes as well as parts of the frontal cortex and cingulate gyrus display severe cognitive and memory deficits. Their knowledge of language structure (at least at early stages of the disease) appears to be relatively intact. The dissociation between language and these functions goes in both directions. Individuals with aphasia show relatively normal cognitive abilities: They can reason, can solve problems (if they do not require language to do so), and have no demonstrable memory deficits; they remember both the past and the present. Thus it appears that language is a functional and neural module.

Nonetheless, recent neuroimaging research has suggested that certain areas of the brain that have been associated with language processing may be recruited as well in other cognitive domains. In this case, the neural area appears to be recruited to perform a particular type of computation or process that is domain general in the sense that it cuts across a number of cognitive domains. Such examples challenge the view that language is a "module" with neural areas specialized only for the processing of components of language or language more generally and suggest that language processing is built upon a set of computational principles shared by other higher cognitive functions.

There are two examples that suggest shared resources in a domain-general manner. Both involve prefrontal areas, including Broca's area (BA44 and BA45) and BA47. It has been suggested that the left inferior frontal gyrus (Broca's area) is recruited in the selection of a representation from among competing representations with the extent of activation modulated by the degree of competition among these representations (Thompson-Schill, 2005; Novick, Trueswell, & Thompson-Schill, 2005; Badre & Wagner, 2007; Miller & Cohen, 2001; Snyder, Feigenson, & Thompson-Schill, 2007; Bilenko, Grindrod, Myers, & Blumstein, 2008).

Competition among competing representations occurs in different levels of language processing. Increased IFG activation has been shown under conditions of competition at the phonological level, for example, selecting a phonetic category such as [t] from acoustically similar phonetic categories such as [d] or [p] (Blumstein et al., 2005; Myers, 2007); at the lexical level, for example, selecting a word such as *can* from the set of words that are similar in their sound shape, such as *pan*, *con*, *cab* (Prabhakaran et al., 2006; Righi, Blumstein, Mertus, & Worden, 2009); at the syntactic level, for example, selecting the appropriate interpretation of an ambiguous sentence such as *put the pear on the paper in the crate* (Novick et al., 2005; Trueswell, Tanenhaus, & Garnsey, 1994); and at the semantic level, for example, selecting the appropriate meaning of an ambiguous word such as *bank*

(river/money) (Bilenko et al. 2008; Mason & Just, 2007). As Thompson-Schill (2005) discusses, increased activation of the IFG also occurs in other cognitive domains under conditions of increased competition including memory, for example, reducing interference in a working memory task (Thompson-Schill et al., 2002), and in visual tasks requiring the maintenance of fixation on a target when another target is in the display (Guitton, Buchtell, & Douglas, 1985). Thus the functional role of the IFG appears to serve as a general cognitive control mechanism that is domain general recruited for both linguistic and nonlinguistic processing.

The prefrontal cortex appears to also play a role in computing categorical representations for both linguistic and nonlinguistic stimuli. The ability to map different sensory stimuli onto a common category plays a critical role in many cognitive domains: We recognize the same face across multiple views, and we perceive the same phonetic category spoken by different speakers. Freedman, Reisenhuber, Poggio, and Miller (2001, 2003) have shown invariant responses in lateral prefrontal cortex to exemplars from a learned visual category (either cats or dogs) in single-cell recordings of monkeys. Similarly, invariant responses in the IFG have been shown for humans who were presented with acoustically different stimuli drawn from the same phonetic category, that is, different exemplars of [t] (Myers, Blumstein, Walsh, & Eliassen, 2009). Thus it appears that the same neural mechanism is recruited in the categorization of both linguistic and nonlinguistic stimuli and that the recruitment of these areas is dictated by the computational requirements underlying categorization rather than the functional domain in which categorization occurs. Taken together, these results suggest that while language may be functionally "special," it draws on at least some neural mechanisms and computational properties that are shared across other cognitive domains. It remains to be seen whether there are other aspects of language that share neural resources across other cognitive domains.

The role of the right hemisphere in language processing

As was discussed earlier, the lesion data support the view that the left hemisphere is dominant for language for most individuals, particularly for right-handers. Patients with right hemisphere lesions are typically not aphasic, and they do not show deficits that can be attributed to linguistic impairments per se. They do, however, show impairments in critical aspects of language communication. In particular, they seem not to be aware of what other speakers know or should know in the communicative process (the theory of mind); they fail to be sensitive to the metaphorical use of language and interpret sentences such as "He has a heavy heart" literally; and they have difficulty in making inferences in sentences (Brownell, Gardner, Prather, & Martino, 1995). Yet, although the neuroimaging literature shows right hemisphere activation in these paralinguistic and discourse tasks, it also often shows bilateral activation in language-processing tasks. In particular, although typically weaker, right hemisphere activation has often been shown in areas homologous to those activated in the left hemisphere in studies exploring phonetic/phonological, lexical, semantic, and syntactic processing (Burton et al., 2000; Bilenko et al., 2008; Rissman, Eliassen, & Blumstein, 2003; Kaan & Swaab, 2002; Mason & Just, 2007). The question then is what role, if any, does the right hemisphere play in linguistic processing. Does it play a functional role? If so, how is its processing integrated with that of the left hemisphere? Is it dominated and hence inhibited by the left hemisphere under normal circumstances, becoming "active" only after left hemisphere injury?

These questions are only starting to be explored. Studies of the neural activation patterns of patients after a stroke and at different stages of language recovery have provided some interesting but conflicting data (for reviews, see Cappa, 2000; Pizzamiglio, Galati, & Committeri, 2001). In some cases, it appears that there is increased activation in the homologous right hemisphere during the recovery process. Other data show increased activation of the right hemisphere in early stages of recovery but activation in perilesional areas for the patients who show the greatest degree of recovery. In fact, increased activation of the right hemisphere is often seen in older subjects in cognitive tasks, raising the question of whether right hemisphere areas reflect a compensatory mechanism or pathological recruitment in the context of language or cognitive capacities during the aging process (Wingfield & Grossman, 2006; Persson et al., 2006; Cabeza, Anderson, Locantore, & McIntosh, 2002).

Agenda for the future

The cognitive neuroscience revolution has enriched our knowledge of the neural systems that underlie language. At the same time, it has raised new questions and challenges that could help to provide a roadmap for future research. First and foremost, localizing areas of the brain that are activated in language processing fails to provide an "explanation" of the functional role those areas play. And it is apparent that no aspect of language activates a single focal neural area. Rather, language and cognitive processing more generally recruit a network of areas. Understanding the functional role of specific neural areas and the neural system in general will occur only when the functional connectivity of these areas is mapped out and when the temporal course of the information flow to and from these areas is delineated. Such research requires the integration of methods that allow for good spatial resolution (e.g., fMRI) with those that allow for good temporal resolution (i.e., MEG, ERPs).

It is generally assumed that activation reflects active involvement of an area in a particular function. This appears not always to be the case. Several studies have shown that a lesion to an area that shows activation in a particular language task does not necessarily result in pathological performance (Price, Mummery, Moore, Frackowiak, & Friston, 1999). Thus the fact that an area is activated in functional neuroimaging does not mean that the area is *necessary* for accomplishing that particular language function. Coupling lesion studies with neuroimaging experiments with normals provides a means of addressing this issue directly. The functional role of an area may be "tested" by examining whether a lesion in an area that is activated in a neuroimaging study results in a functional deficit. Alternatively a lesioned area giving rise to a particular deficit should show activation in normal individuals using neuroimaging methods, assuming that comparable language stimuli and experimental tasks are used.

Finally, it is not at all clear whether a component of the grammar can ever be studied and identified independent of a particular task. After all, the function of language is to communicate and the use of language whether for speaking or understanding is a goal-directed behavior. Does it make sense, then, to talk about sound structure, lexical structure, semantic structure, or syntactic structure independent of its use? That is, accessing sound structure to identify a sound segment is different from accessing sound structure to compare to another sound or mapping a sound on to lexical form (cf. Poeppel, 1996). Yet each of these tasks recruits the "linguistic component" as well as a set of other cognitive and executive resources needed to accomplish the specific goal. In the end, it would seem that understanding the neural basis of language requires the study of language in action. If this is the case, then it less clear that there will ever be a true dissociation between a linguistic component per se and the processes and mechanisms underlying its use.

ACKNOWLEDGMENTS This research was supported in part by NIH Grants RO1 DC006220 and RO1 DC00314 from the National Institute on Deafness and Other Communication Disorders. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

REFERENCES

BADRE, D., & WAGNER, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(3), 2883–2901.

BATES, E., FRIEDERICI, A., & WULFECK, B. (1987). Comprehension in aphasia: A cross-linguistic study. *Brain Lang.*, *32*(1), 19–67.

BAUM, S., BLUMSTEIN, S., NAESER, M., & PALUMBO, C. (1990). Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain Lang.*, *39*(1), 33–56.

BILENKO, N., GRINDROD, C., MYERS, E. B., & BLUMSTEIN, S. E. (2008). Neural correlates of semantic competition during processing of ambiguous words. *J. Cogn. Neurosci.*, *21*(5), 960–975.

BLUMSTEIN, S. E. (2000). Deficits of speech production and speech perception in aphasia. In R. Berndt (Ed.), *Handbook of neuropsychology* (2nd ed., Vol. 2, pp. 95–113). Amsterdam: Elsevier Science.

BLUMSTEIN, S. E., MYERS, E. B., & RISSMAN, J. (2005). The perception of voice-onset time: An fMRI investigation of phonetic category structure. *J. Cogn. Neurosci.*, *17*(9), 1353–1366.

BROWNELL, H., GARDNER, H., PRATHER, P., & MARTINO, G. (1995). Language, communication, and the right hemisphere. In H. S. Kirshner (Ed.), *Handbook of neurological speech and language disorders* (pp. 325–349). New York: Marcel Dekker.

BUCHSBAUM, B., HICKOK, G., & HUMPHRIES, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.*, *25*(5), 663–678.

BUCKNER, R. L., RAICHLE, M. E., & PETERSEN, S. E. (1995). Dissociation of human prefrontal cortical areas across different speech production tasks and gender groups. *J. Neurophysiol.*, *74*, 2163–2173.

BURTON, M. W. (2001). The role of inferior frontal cortex in phonological processing. *Cogn. Sci.*, *25*(5), 695–709.

BURTON, M. W. (2009). Understanding the role of prefrontal cortex in phonological processing. *Clin. Linguistics Phonetics*, *23*(3), 180–195.

BURTON, M. W., SMALL, S. L., & BLUMSTEIN, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *J. Cogn. Neurosci.*, *12*(4), 679–690.

CABEZA, R., ANDERSON, N. D., LOCANTORE, J. K., & MCINTOSH, A. R. (2002). Aging gracefully: Compensatory brain activity in high-performing older adults. *NeuroImage*, *17*(3), 1394–1402.

CAPLAN, D. (1992). *Language: Structure, processing, and disorders.* Cambridge, MA: MIT Press.

CAPLAN, D. (1994). Language and the brain. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1023–1053). New York: Academic Press.

CAPPA, S. F. (2000). Neuroimaging of recovery from aphasia. *Neuropsychol. Rehabil.*, *10*(3), 365–376.

CAPPA, S. F., & PERANI, D. (2006). Broca's areas and lexical-semantic processing. In Y. Grodzinsky & K. Amunts (Eds.), *Broca's region* (pp. 187–195). Oxford, UK: Oxford University Press.

DRONKERS, N. F., WILKINS, D. P., VAN VALIN, R. D., JR., REDFERN, B. B., & JAEGER, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1–2), 145–177.

FAROQI-SHAH Y., & THOMPSON C. K. (2003). Effect of lexical cues on the production of active and passive sentences in Broca's and Wernicke's aphasia. *Brain Lang.*, *85*(3), 409–426.

FIEZ, J. A. (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. *Hum. Brain Mapping*, *5*, 79–83.

FODOR, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

FREEDMAN, D. J., RIESENHUBER, M., POGGIO, T., & MILLER, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*(5502), 312–316.

FREEDMAN, D. J., RIESENHUBER, M., POGGIO, T., & MILLER, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.*, *23*(12), 5235–5246.

GOODGLASS, H. (1993). *Understanding aphasia.* New York: Academic Press.

GRODZINSKY, Y. (1990). *Theoretical perspectives on language deficits.* Cambridge, MA: MIT Press.

Grodzinsky, Y., & Amunts, K. (Eds.). (2006). *Broca's region*. Oxford, UK: Oxford University Press.

Grodzinsky, Y., Shapiro, L., & Swinney, D. (2000). *Language and the brain: Representation and process*. New York: Academic Press.

Guitton, D., Buchtel, H. A., & Douglas, R. M. (1985). Frontal lobe lesions in man cause difficulties in suppressing reflexive glances and in generating goal-directed saccades. *Exp. Brain Res.*, *58*(3), 455–472.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.*, *28*(10), 2539–2550.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *J. Cogn. Neurosci.*, *15*(5), 673–682.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.*, *4*(4), 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, *274*(5284), 114–116.

Kaan, E., & Swaab, T. (2002). The brain circuitry of syntactic comprehension. *Trends Cogn. Sci.*, *6*(8), 350–356.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Martin, R. C., Vuong, L. C., & Crowther, J. E. (2007). Sentence-level deficits in aphasia. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*. Oxford, UK: Oxford University Press.

Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Res.*, *1146*, 115–127.

Milberg, W., Blumstein, S. E., Giovanello, S. S., & Misiurski, C. (2003). Summation priming in aphasia: Evidence for alterations in semantic integration and activation. *Brain Cogn.*, *51*(1), 31–47.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex. *Annu. Rev. Neurosci.*, *24*, 167–202.

Myers, E. B. (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia*, *45*(7), 1463–1473.

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol. Sci.*, in press.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension *Cogn. Affective Behav. Neurosci.*, *5*(3), 263–281.

Paulesu, E., Frith, C. D., & Frackowiak, R. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, *25*(362), 342–345.

Persson J., Nyberg, L., Lind, J., Larsson, A., Nilsson, L. G., Ingvar, M., & Buckner, R. L. (2006). Structure-function correlates of cognitive decline in aging. *Cereb. Cortex*, *16*(7), 907–915.

Piñango, M. M., & Zurif, E. B. (2001). Semantic operations in aphasic comprehension: Implications for the cortical organization of language. *Brain Lang.*, *79*(2), 297–308.

Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: William Morrow.

Pizzamiglio, L., Galati, G., & Committeri, G. (2001). The contribution of functional neuroimaging to recovery after brain damage: A review. *Cortex*, *37*(1), 11–31.

Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain Lang.*, *55*(3), 317–351.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cogn. Sci.*, *25*(5), 679–693.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrielli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage*, *10*(1), 15–35.

Prabhakaran, R., Blumstein, S. E., Myers, E. B., & Hutchison, E. (2006). An event-related fMRI investigation of phonological-lexical competition, *Neuropsychologia*, *44*(12), 2209–2221.

Price, C. J., Mummery, C. J., Moore, C. J., Frackowiak, R. S. J., & Friston, K. J. (1999). Delineating necessary and sufficient neural systems with functional imaging studies of neuropsychological patients. *J. Cogn. Neurosci.*, *11*(4), 371–382.

Righi, G., Blumstein, S. E., Mertus, J., & Worden, M. S. (2009). Neural systems underlying lexical competition: An eyetracking and fMRI study. *J. Cogn. Neurosci.* [Epub ahead of print. doi: 10.1162/jocn.2009.21200.]

Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related fMRI investigation of implicit semantic priming. *J. Cogn. Neurosci.*, *15*(8), 1160–1175.

Scott, S. K., & Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, *92*(1–2), 13–45.

Shelton, J. R., & Caramazza, A. (1999). Deficits in lexical and semantic processing: Implications for models of normal language. *Psychon. Bull. Rev.*, *6*(1), 5–27.

Snyder, H. R., Feigenson, K., & Thompson-Schill, S. L. (2007). Prefrontal cortical response to conflict during semantic and phonological tasks. *J. Cogn. Neurosci.*, *19*(5), 761–775.

Thompson-Schill, S. L. (2005). Dissecting the language organ: A new look at the role of Broca's area in language processing. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 173–189). Mahwah, NJ: Lawrence Erlbaum Associates.

Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D'Esposito, M., Kan, I. P., Knight, R. T., & Swick, D. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cogn. Affective Behav. Neurosci.*, *2*(2), 109–120.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *J. Mem. Lang.*, *33*(3), 285–318.

Wingfield, A., & Grossman, M. (2006). Language and the aging brain: Patterns of neural compensation revealed by functional brain imaging. *J. Neurophysiol.*, *96*(6), 2830–2839.

Yee, E., Blumstein, S. E., & Sedivy, J. C. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: Evidence from eye movements. *J. Cogn. Neurosci.*, *20*(4), 592–612.

87 Why the Imagery Debate Won't Go Away

STEPHEN M. KOSSLYN, WILLIAM L. THOMPSON, AND GIORGIO GANIS

ABSTRACT In this chapter, we summarize the theoretical and empirical history of the imagery debate. We argue that the debate about the nature of mental imagery has lasted so long in part because it taps into much more general issues, which bear on the nature of theories of cognitive functions. The imagery debate has focused on whether depictive representations are used in cognition, but the depictive and antidepictive camps have often failed to engage each other, in part because they have asked questions at different levels of analysis. Depictive theories have focused on the nature of internal representation and processing, which has resulted in their embracing computational and neuroscientific evidence to explain data from experiments and to produce empirical predictions. In contrast, antidepictive theories have tended to focus on competence, developing abstract explanatory principles couched in the language of formal logic, with little attention to processing per se. Rather than being in direct conflict, the two sorts of theories often can complement each other.

The so-called imagery debate began in 1973, with publication of Zenon Pylyshyn's paper "What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery." Pylyshyn argued that just as the heat emitted by a light bulb plays no functional role in the process of reading, the pictorial aspects of mental imagery play no role in information processing. Kosslyn and Pomerantz (1977) replied to this article, and the debate was off and running. The imagery debate is not about whether people report experiencing picture-like mental images (they do), nor is it about whether visual information is stored and used in memory and reasoning (yes, it is); moreover, the debate is not about whether mental imagery shares mechanisms with perception (yes, it does). Rather than addressing whether visual *content* can be stored in memory and used in reasoning, the debate is about the *format* of the representation—the nature of the code itself. Whereas some of us argue that visual mental images are called *images* because they in fact share structural properties with actual pictures (in particular, space in a representation is used to depict the layout of surfaces in actual space), others argue that only a single kind of representation is used in all of cognition and that this representation uses language-like symbols and is best understood in the context of the characteristics of formal logic. This is the "propositional" view of mental imagery.

Uncounted salvos and exchanges followed, and long after most colleagues have lost interest, the debate still smolders. Why won't the imagery debate go away, and why should anybody care whether it does or doesn't? In this brief chapter, we expand on our earlier views (Kosslyn, Thompson & Ganis, 2006) and explain how the lessons of the imagery debate apply more broadly to cognitive psychology, cognitive science, and cognitive neuroscience as a whole.

What is a theory of cognition?

The imagery debate has persisted in part because the two camps have sought different types of theories, which have different purposes. The issues that have been revealed apply broadly.

COMPETENCE VERSUS PROCESSING When considering theories of cognitive processing in general and imagery in particular, we must emphasize a crucial distinction between two classes of theories. On the one hand, *theories of competence* specify principles that describe a system; the system itself does not actually follow the rules specified by these principles any more than a seal solves differential equations when it hunches its neck to catch a ball on its nose or the planets follow the laws of motion when they wheel around the sun. On the other hand, *theories of processing* specify mechanisms that underlie performance; in so doing, they indicate how information (including specifications of rules) is represented and processed (including by processes that actually follow rules) in a system in real time (cf. Chomsky, 1965).

Not only are the two types of theories designed to answer different types of questions, but the answers themselves are qualitatively distinct. Depictive theorists, who propose that visual mental images rely on depictive, picture-like representations, have sought to explain the details of empirical findings by appeal to mechanisms, which requires them to specify the nature of representations and the processes that operate on these representations. In contrast, antidepictive

STEPHEN M. KOSSLYN and WILLIAM L. THOMPSON Harvard University, Cambridge, Massachusetts
GIORGIO GANIS Massachusetts General Hospital, Charlestown, Massachusetts

theorists, who posit that a single language of thought underlies all forms of cognition, have more recently tried to explain previous findings with simple, general principles. For example, antidepictive theorists seem satisfied that they have explained a finding by appealing to "tacit knowledge" of how perception works; according to this view, such knowledge is unconscious and can guide behavior in laboratory experiments. For instance, tacit knowledge is taken to explain the increase in time to scan farther distances over an object in an image. According to this view, during imagery, people unconsciously simulate what they believe would happen in perception, which leads them to wait a little longer to respond when they should have traversed greater distances while scanning.

The initial phases of the debate pitted depictive and propositional theories against each other, which was appropriate, since both began as theories of processing. However, with the advent of tacit knowledge accounts (Pylyshyn, 1981), the debate shifted. The alternative to depictive theories was no longer another mechanistic theory, but—as far as we can tell—a theory of competence (e.g., of the sort that Pylyshyn [2002] believes has been successful in characterizing language). That is, the theory aims to describe what people know (largely unconsciously) and ways in which that knowledge can be used, without characterizing the mechanisms that give rise to performance.

The two sorts of theories can, of course, go hand in hand. A theory of competence can provide guidelines that must be met by theories of processing. The two kinds of theories are not exclusive, which might be one reason why the imagery debate has proven so slippery: The theorists are talking past each other.

Formal Systems Another reason that the imagery debate has persisted lies in the different conceptions of how a theory should be stated. The nondepictive camp has asserted that theories must be stated by using formalisms such as logic or mathematics. But is this necessarily so? Darwin's answer would be clear, and the success of his theory of evolution by natural selection serves as a model for the utility of nonformal theories.

Pylyshyn and his colleagues urge us to consider the most general constraints or boundary conditions that have to be met by a system of imagistic reasoning. For example, Fodor and Pylyshyn (1988) have proposed such constraints as productivity, compositionality, and systematicity. In addition, Pylyshyn (1978) has proposed some formal constraints on theories of imagery; for example, they must represent token individuals and visual rather than abstract properties, and they must not encode properties of sets, such as cardinality, or represent negative statements. Although plausible, these very abstract ideas do not go very far in explaining the data, such as increases in time as objects in images are rotated by greater amounts (Shepard & Cooper, 1982). In other words, these formal constraints are not very illuminating from the point of view of a theory of processing.

Computational Models Seen from another perspective, the debate has also continued because the depictive camp has sought to develop models that do not make close contact with the sorts of theories developed by most antidepictive theorists. Specifically, the depictive theorists have relied on theoretical constructs borrowed from computer science, arguing that computer simulation models provide one way to begin to specify crisply concepts that are exceedingly slippery and difficult to nail down using natural language alone. Models, by their nature, are not fully specified (Hesse, 1963).

In contrast to this approach, Pylyshyn (2002, 2003) not only argues that the depictive theory is incorrect, but also goes so far as to suggest that the issue of the format of image representations is actually irrelevant. According to this view, pursuing knowledge about the format of images tells us nothing truly interesting about imagery. If Pylyshyn is (implicitly) focusing on a formal theory of competence, we can make sense of this claim. In this case, he would seek an abstract characterization of what people know, not how information is represented and processed when one actually performs specific tasks.

If one seeks a mechanistic theory, then it is difficult to understand the claim that characterizing the format of the representation is irrelevant. Rather, such a goal focuses us on a fundamental question about the nature of information processing: Is there a single "language of thought"? The answer to this question has wide-ranging implications for studies of many faculties. On the one hand, if it turns out that there really is only a single type of internal representation, this fact would place very strong constraints on theories of all forms of cognition. On the other hand, if it turns out that there are different types of representations, this opens the door to theories of specialized mechanisms for specialized tasks. To address the question of the format of representations in imagery, one needs to specify a theory of processing mechanisms.

Motivating a Theory Pylyshyn (again taking him as the prototypical antidepictive theorist) considers it a major drawback if a theory is "speculative," and he simply dismisses such theories out of hand; his view seems to be that a theory must be considered as "speculative" if any of its details are not motivated by proven facts. In contrast, we (taking ourselves as typical depictive theorists) maintain that there is a sense in which *any* theory is "speculative" in this way. We know of no case in science in which an "adequate theory" (from Pylyshyn's point of view) was in place before properties of the subject matter were well characterized. In

fact, it often goes the other way around: Theories are constructed by an act of bootstrapping, general principles being abstracted out of individual discoveries, which in turn were guided by earlier theories (e.g., see Fodor, 2003). It is not possible to motivate every detail of a theory in advance; part of what makes a theory a "theory" (as opposed to a description of fact) is that it is a stab into the unknown.

Much of the motivation for the formal constraints imposed by antidepictive theorists appears to be based on introspection and logical considerations. Therefore it is not surprising that some of these claims are debatable. For example, Barsalou (1999) challenges the idea that images must represent only token individuals; instead, he offers detailed arguments that perceptual representations can perform all the functions often assumed to require symbolic or amodal representations. Again, the camps appear to be barking up different trees.

What Theories Are For Another point of fracture focuses on what work we want theories to do. Depictive theorists stress the importance of generating and testing new predictions, whereas antidepictive theorists are usually content to try to explain previous findings. Depictive theorists have stressed that their theories have led to an avalanche of new empirical discoveries (e.g., Kosslyn, 1980, 1994; Kosslyn, Ganis & Thompson, 2001; Kosslyn & Thompson, 2003; Kosslyn et al., 2006; Thompson & Kosslyn, 2000), whereas it is not clear that antidepictive theories have led to many new discoveries. The antidepictive camp downplays the role of depictive theories in the empirical enterprise for the reasons noted above, claiming that these theories are not really theories at all.

No wonder the imagery debate has ground on for so long!

Relevance of the brain

We were moved to write this chapter now because the debate has resurged in recent years. This uptick in discussion is a direct result of a change in research strategy. Specifically, the depictive camp has come to focus on neural evidence. Anderson (1978) showed that certain classes of behavioral evidence could be easily accounted for by either depictive or nondepictive explanations. To resolve this fundamental ambiguity, he proposed that additional sources of evidence, such as neural data, should be taken into account. However, the currently popular antidepictive approach rejects the very idea that facts about the brain can be used to discover the format of an internal representation.

The Role of Neuroanatomy As an example of the lack of utility of using facts about neuroanatomy to constrain theorizing about cognitive function, Pylyshyn points out that knowing about the projections from the retina to the brain has led to fruitless debates about why we don't see the world upside down. In sharp contrast, the depictive camp emphasizes that studying the wiring of the visual system has led to enormous progress in our understanding of how vision works (e.g., Desimone & Ungerleider, 1989; Felleman & Van Essen, 1991; Hubel, 1988). For instance, the fact that object-properties processing and spatial-properties processing are accomplished by separate neural systems allows us to understand why we sometimes erroneously conflate shape and location information. Similarly, by observing the relative amount of cortex devoted to the foveal versus peripheral areas of the retina, we gain insight into the nature of visual acuity. Moreover, Anderson (1978) used structure/process tradeoffs to show how depictive theories of behavioral findings could be converted to antidepictive theories. Such tradeoffs require being able to specify the nature of processing to fit the properties of representations. But once we know certain facts about the brain, such tradeoffs can no longer be made—we can no longer alter properties of representations and processes willy-nilly to perform such tradeoffs.

Why do the two camps have opposite evaluations of the potential utility of neuroscientific data? One reason might be that no hint of a mapping to neural mechanisms has been found for theories that posit only tacit knowledge or only propositional representations. This is not surprising, given the abstract level of analysis that the antidepictive theories have typically taken.

Types of Data In addition, the two camps have different assumptions about what sorts of data bear on the issues. According to some antidepictive theorists (e.g., Dennett, 2002; Pylyshyn, 2002), *only* behavioral evidence can reveal the nature of information processing. This view sharply contrasts with that of the depictive theorists, who note that researchers can learn an enormous amount about the types and course of information processing by tracking the spatial and temporal properties of activity in the brain itself, (e.g., Ganis, Thompson, & Kosslyn, 2004; Ganis & Schendan, 2008). From the perspective of depictive theories, the approach advocated by the opposing camp throws away valuable tools and evidence that can help us to understand the system that underlies all information processing. According to this view, it is as if during the early stages of research on the building blocks of life, researchers had decided to throw away X-ray crystallography data by assuming that it had nothing to do with understanding "life." (But without this tool, they would not have discovered DNA!)

The Value of Studying Internal Events Some antidepictive theorists—perhaps as an outgrowth of their distrust of introspection—have discounted the value of studying internal events that do not lead to overt, observable behavior.

Again in contrast, the view that most depictive theorists espouse is that by studying the brain, we can discover the nature of processing that never leads to an overt response. This divergence in viewpoints bears on issues much broader than the nature of mental imagery. To take a simple example, it has long been known that after extinction of a classically conditioned stimulus-response pair, animals later can be conditioned again more easily than they were at the outset. But why is this so? Is it because there is still some residual learning that was not affected by extinction and therefore not all neural connections need to be reestablished? Is it because extinction replaces the conditioned response with a new one and relearning consists of removing the new response? As it happens, LeDoux and colleagues (reviewed in LeDoux, 1996) have shown that extinction of classically conditioned responses relies on the frontal lobe's suppressing reflexive connections (mediated by the amygdala). That is, extinction does not obliterate or overwrite the conditioned memory; it simply keeps it in check, but that memory is always there, lurking in the background. Observing behavior alone would never tell us this because, by all overt appearances, the behavior has been eliminated and then can be relearned more easily—but without access to the neuroscientific data, we might never have learned why.

A Special Role for Neurological Data? Another point of divergence is that some in the antidepictive camp assert that biological or neuroscience-based data do not have a special role in evaluating theories of information processing. This claim is ironic, given that Pylyshyn's (1981) shift to focusing on the alternative interpretations of key imagery findings motivated us to turn to the brain. Specifically, Pylyshyn claimed that the results of classic imagery experiments, such as mental rotation or mental scanning (for review, see Kosslyn et al., 2006), do not directly reflect the nature of internal mechanisms. Rather, as was noted earlier, Pylyshyn speculates that such results are produced on the basis of participants' tacit knowledge about the corresponding perceptual phenomena—and that knowledge leads the participants (unconsciously) to try to mimic what they think they would have done in the corresponding perceptual situation.

In this context, neuroanatomical and neurophysiological facts do play a special role. To the extent to which we shift from measuring overt behavior to measuring neural events, we take tacit knowledge out of the picture. For example, the brain has many visual areas that are topographically organized; the structure of activation across the retina is physically laid out (albeit with various distortions) across the surface of the cerebral cortex. Many studies (but not all; see Kosslyn & Thompson, 2003) have now shown not only that these areas are activated when one visualizes an object with high resolution, but also that spatial properties of the object (such as its orientation) are directly reflected by the pattern of activation. These representations are not language-like and are not abstract descriptions; they are literally depictions: Activation in specific locations of space on cortex specifies objects in space in the real world. No amount of tacit knowledge can explain them away.

In addition, facts about the brain introduce constraints that prevent theories from being arbitrarily modified to fit data. For example, any theory of visual processing must respect the fact that these areas are topographically organized. Similarly, many additional facts about the brain constrain theories of information processing, such as the fact that object properties and spatial properties are processed by largely separate neural systems, that color and motion are processed by partially distinct systems, that different systems subserve different sorts of working memory, and so on and so forth. Such constraints affect all theories.

Some general lessons from the imagery debate

Why should anybody care about the imagery debate? Because the debate strikes to the heart of issues regarding what the study of the mind should be. Can we characterize the nature of internal representations and is it even worth trying to do so? Should we try to characterize mental function at a very abstract level or at an information processing level? Or should we do both? Should theories be stated formally, or can they do useful work when cast as partially specified models?

The imagery debate and the ways in which the arguments on either side have failed fully to engage each other have brought into sharp relief three general lessons. First, we need to be clear on the goals of theorizing. What do we want to use theories to do? In cognitive neuroscience, we are trying to characterize a system, and in so doing, we need to characterize the individual components and how they interact. This effort requires analysis at multiple levels. Not only must we understand what individual neurons are doing and how neurons affect each other, but we also need to understand how large ensembles of neurons come into play and interact, and we need to understand what these events are accomplishing. As Marr (1982) stressed, we need to specify what is being computed as well as how the computation is taking place. In much of contemporary cognitive neuroscience, we have set aside the "what" part and have focused solely on the "how."

Second, ultimately, we will need to engage in formal theorizing. Cognitive science, with its emphasis on computational models, might not have had a sufficient influence on the empirical work in contemporary cognitive neuroscience. But these sorts of models are likely to be useful primarily during a transitional phase, before truly rigorous theorizing underlies research in cognitive neuroscience. However, in

the early years of the 21st century, it is not clear what will be the most appropriate formalisms for characterizing the brain; it is possible that the most useful formal vocabulary, capable of capturing the complex dynamics of the brain, has not even been invented yet. In the meantime, although it is reasonable to use formalisms that are available, we must avoid being dogmatic about them. We must avoid the "man-with-hammer" phenomenon: To the man who has only a hammer, the whole world looks like a nail. To the person with a particular formalism, every phenomenon may look like fair game—but it might not be.

Third, we need to be clear about the best ways to evaluate our theories. Many metrics are possible, as the imagery debate has illustrated. Surely, clarity and elegance are to be valued, but so is predictive power.

In short, cognitive neuroscience must fully embrace the idea that neural activity should be characterized as computation, and one aspect of understanding a computational system is understanding the sorts of representations that can be used. Moreover, characterizing the format of the representations used in cognition is one part of this challenge. When we visualize shapes with high resolution, either a depictive representation is employed or it is not. If cognitive neuroscience is a science, we should be able to resolve this issue.

ACKNOWLEDGMENTS Preparation of this chapter was supported by grants R01 MH060734 from the National Institutes of Health and REC-0411725 from the National Science Foundation. We thank the Oxford University Press for allowing us to adapt part of this chapter from portions of chapter 6 of Kosslyn, Thompson, and Ganis (2006). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institutes of Health or the National Science Foundation.

REFERENCES

ANDERSON, J. R. (1978). Arguments concerning representations for mental imagery. *Psychol. Rev.*, *85*, 249–277.

BARSALOU, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.*, *22*, 577–660.

CHOMSKY, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

DENNETT, D. C. (2002). Does your brain use the images in it, and if so, how? *Behav. Brain Sci.*, *25*, 189–190.

DESIMONE, R., & UNGERLEIDER, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (pp. 267–299). Amsterdam: Elsevier.

FELLEMAN, D. J., & VAN ESSEN, D. C. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cereb. Cortex*, *1*, 1–47.

FODOR J. A. (2003). *Hume variations*. Oxford, UK: Clarendon Press.

FODOR J. A., & PYLYSHYN, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

GANIS, G., & SCHENDAN, H. E. (2008). Visual mental imagery and perception produce opposite adaptation effects on early brain potentials. *NeuroImage*, *42*, 1714–1727.

GANIS, G., THOMPSON, W. L., & KOSSLYN, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Brain Res. Cogn. Brain Res.*, *2*, 226–241.

HESSE, M. (1963). *Models and analogies in science*. London: Sheed & Ward.

HUBEL, D. H. (1988). *Eye, brain and vision*. New York: W. H. Freeman.

KOSSLYN, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

KOSSLYN, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

KOSSLYN, S. M., GANIS, G., & THOMPSON, W. L. (2001). Neural foundations of imagery. *Nature Rev. Neurosci.*, *2*, 635–642.

KOSSLYN, S. M., & POMERANTZ, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cogn. Psychol.*, *9*, 52–76.

KOSSLYN, S. M., & THOMPSON, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychol. Bull.*, *129*, 723–746.

KOSSLYN, S. M., THOMPSON, W. L., & GANIS, G. (2006). *The case for mental imagery*. Oxford, UK: Oxford University Press.

LEDOUX, J. (1996). *The emotional brain*. New York: Simon & Schuster.

MARR, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman.

PYLYSHYN, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychol. Bull.*, *80*, 1–24.

PYLYSHYN, Z. W. (1978). Imagery and artificial intelligence. In C. W. Savage (Ed.), *Perception and cognition: Issues in the foundations of psychology. Minnesota studies in the philosophy of science* (Vol. 9, pp. 19–55). Minneapolis: University of Minnesota Press.

PYLYSHYN, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychol. Rev.*, *87*, 16–45.

PYLYSHYN, Z. W. (2002). Mental imagery: In search of a theory. *Behav. Brain Sci.*, *25*, 157–238.

PYLYSHYN, Z. W. (2003). Return of the mental image: Are there pictures in the brain? *Trends Cogn. Sci.*, *7*, 113–118.

SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

THOMPSON, W. L., & KOSSLYN, S. M. (2000). Neural systems activated during visual mental imagery: A review and meta-analyses. In A. W. Toga & J. C. Mazziotta (Eds.), *Brain mapping: The systems* (pp. 535–560). San Diego, CA: Academic Press.

88 Looking Toward the Future: Perspectives on Examining the Architecture and Function of the Human Brain as a Complex System

MICHAEL S. GAZZANIGA, KARL W. DORON, AND CHADD M. FUNK

ABSTRACT With the arrival of increasingly sophisticated methodologies, cognitive neuroscience has quickly matured into a remarkably prolific field. However, amid such progress, it is possible to lose sight of the overarching purpose of the field, which is to promote an understanding of how the brain creates the marvels of the mind. We argue that viewing the brain as a complex system is a useful way of regaining this perspective. Here, we consider fundamental characteristics of brain organization and function in hopes of finding analogous complex systems that might yield novel insight into how cognitive function and subject experience emerge from activity in distributed regions of the brain.

The aim of cognitive neuroscience is to advance our understanding of the organization and function of the human brain and, ultimately, to solve the mystery of how the human brain creates the human mind. Moving toward these ends occasionally requires dismissing obsolete theoretical constructs and taking new perspectives in order to devise the theories that will succeed them. This chapter is an effort to do both. We first consider the modular organization of the brain; we then address why it renders some descriptions of brain function—namely, those that appeal to a homunculus-like entity—inadequate. Though slowly fading, the notion that a homunculus is embedded somewhere inside the depths of the brain persists, subtly masked by terms such as *executive function* or *cognitive control*. This reliance upon a central executive or neural hub that receives feedforward input and distributes attention, induces actions, and carries the cognitive burden is likely a remnant of outdated views of the brain that assume that perception, cognition, and action are discrete functions of the human brain that always occur in a prescribed hierarchical sequence.

Accumulating evidence has revealed that the brain is organized into modules that operate in parallel rather than in any simple progression, leaving no single neural apex to act as an executive. Cognitive neuroscience must move forward without a central hierarchy, even though there are significant challenges inherent in forging a new path. We would like to add our voice to those of the people who view the human brain as a complex system. We wish to promote the application of this type of theoretical inquiry within the field of cognitive neuroscience and further demonstrate that, by doing so, we may dispose of the fabled homunculus. Viewing the brain as a complex system provides a worthwhile perspective from which to understand how the interconnected modular architecture of the brain generates the many emergent properties studied by cognitive neuroscience.

In what follows, we describe several complex systems and draw parallels and contrasts between these systems and human brain activity at the global level, ultimately revealing the brain as a complex and dynamic system. In this view, emergent properties of the organizational and functional structure of the brain are the result of the interaction of neuronal assemblies, thus eliminating the need for a central controller.

The modular hypothesis as successor to traditional views

One of the most radical truths to emerge from decades of split-brain research is that the human brain is not an

MICHAEL S. GAZZANIGA SAGE Center for the Study of Mind and Department of Psychology, University of California, Santa Barbara, California

KARL W. DORON and CHADD M. FUNK Department of Psychology, University of California, Santa Barbara, California

all-purpose computing device but instead is organized in modular fashion, consisting of specialized circuits that have been sculpted by evolution to efficiently perform specific functions (Gazzaniga, 1989; Baynes, Eliassen, Lutsep, & Gazzaniga, 1998). This observation has been complemented by studies of patients with various other neurological disorders (Cooney & Gazzaniga, 2003) and by an ever-growing body of functional brain imaging literature. This architecture accounts for the unexpectedly enormous amount of distributed nonconscious processing that goes on from moment to moment, since parallel activity across modules can facilitate vast amounts of simultaneous activity (Volpe, Ledoux, & Gazzaniga, 1979). Functional modules typically consist of hierarchal processing, involve long-distance interactions between distributed neural regions, and also interact with or share neural real estate with numerous other modules. However, the attribute of this organization that is most relevant for the discussion that follows is the lack of an intermodular hierarchy. If correct, and accumulating evidence is quickly reducing remaining doubts, this theory challenges certain prominent undercurrents in psychology and cognitive science that have guided research and theory for nearly a half-century.

Cognitive science, a major precursor to cognitive neuroscience, has come a long way since triumphing over the behavioralist movement in the middle of the 20th century (Miller, 2003). In the 1950s, after decades focused entirely on measurable behaviors rather than the cognitive events that triggered them, the idea that the brain did something more than act as a mere organ for learning responses to input from the world took hold, never again to relinquish its grip. This revolution enabled extensive investigation of the function of the brain and eventually gave birth to modern cognitive neuroscience. Though progress has led to constant revision of specific theories and models of various brain functions, certain fundamental principles have endured. Central among these is the idea that cognition is purposefully wedged between sensation and action and that sharp boundaries isolate the components of this tripartite conception. However, implicit in this model is a hierarchy in which all the information derived from perception is processed and eventually passed on to some decision-making area at the apex that selects a motor response. This theoretical neural pinnacle, which has inevitably remained elusive, is tantamount to the fabled homunculus, which receives information from all areas of the brain and pulls the strings necessary to cause actions.

The modular hypothesis challenges the orderly sequence of perception, cognition, and action and, in doing so, destroys the chamber occupied by the homunculus. As we acquire a deeper understanding of the function of the brain, it is becoming clear that many functions defy simple classification as perception, cognition, or action. Indeed, the view of the motor system surfacing in this volume is one in which its components play an important role in many functions that have traditionally been thought of as cognitive, such as theory of mind (chapter 43 in this volume) and intention formation (chapter 83 in this volume). Also, a separate line of research indicates that regions necessary for control of eye movements also play a role in attention (Wardak, Ibos, Duhamel, & Olivier, 2006).

Meanwhile, areas associated with perception may also be involved in processes that have traditionally been deemed cognitive. For instance, cortical regions that support perception may also provide the basis for object recognition (chapter 32 in this volume). Also, bottom-up forms of attention, based on salience, indicate that enhanced processing could be promoted by intramodular activity (chapter 13 in this volume). These are just a few of the many examples of phenomena that violate conventional categorization. Some might debate the above statements, but that is precisely the point. We believe that time would be better spent on accepting and further exploring the diverse nature of brain function as a consequence of parallel processing in which modules influence activity in other modules in myriad ways, leading to subjective experience and behavioral output, than on arguing about how to classify hybrid functions.

Furthermore, it is quite clear that the organization of the brain cannot accommodate a homunculus or any similar entity. This should not come as a surprise to philosophers and psychologists alike, who have already banished the homunculus from the brain on other grounds, such as the principle of infinite regression (surely the homunculus needs a homunculus; see Niesser, 1967). However, cognitive neuroscientists often still cling to a subtle extension of this doctrine that operates under the guise of "executive function" or the "top" in "top-down." There are a number of reasons why this might be the case. It is, for instance, tempting to think that the well-established limited capacities of attentional processing imply a central executive capable of only serial cognitive function (Holtzman & Gazzaniga, 1982; Dux, Ivanoff, Asplund, & Marois, 2006). But a bottleneck does not mandate a final common cognitive pathway for every type of upper-level neural processing. The far more likely explanation of limited cognitive resources is that certain tasks require the same modules that have limited capacity. This explains how other tasks, which might require separate resources, could be performed in the absence of attention. For instance, subjects can make discriminations about whether or not an animal appeared in pictures shown peripheral to a demanding central task (Li, VanRullen, Koch, & Perona, 2002).

The persistence of the homunculus likely reflects the common forceful intuition of being a unified agent with a strong sense of self. The mystery of this persuasive intuition certainly must be addressed (Crick & Koch, 2000, 2003), but the solution will not depend on the existence of a traditional

homunculus. Nowhere is this more evident than in chapters scattered throughout this volume. For instance, research on networks for self-regulation of emotional responses (chapter 66 in this volume) and cognitive retrieval of episodic memory (chapter 48 in this volume) indicates that even the seemingly unitary notion of cognitive control or executive function is implemented by distributed neural regions with distinct functions. Others have provided evidence for distinct networks involved in task control that act on different timescales and by different mechanisms (Dosenbach et al., 2007). Each of these lines of research further attenuates the potential repertoire of the homunculus, revealing that cognitive neuroscience no longer has any need for this theoretical crutch.

Moving toward a comprehensive theory of brain function

As the field replaces these notions with a modular architecture in which modules act in parallel to perform varied functions without guidance of an all-powerful central executive, it is imperative that these observations are fit within an explanatory framework that facilitates further research and comprehension. This is the new challenge faced by the field of cognitive neuroscience, and it requires a somewhat different approach from the one that has brought us this far. The conclusions discussed above are largely the result of careful scrutiny of specific modules or components of modules in order to elucidate their anatomical basis and function, an undeniably worthwhile and fruitful approach that attempts to connect neural cause with functional effect. However, the ultimate goal of cognitive neuroscience is to understand how the many modules of the brain interact in a way that explains complex human experience and behavior. The conscious experience generated by the human brain is both integrated and coherent, and our many capabilities related to perception, cognition, and action rarely act in noticeable isolation. Clearly, then, integration is essential and has rightfully been at the center of recent quantitative theoretical approaches to brain function (Tononi, Sporns, & Edelman, 1994; Seth, Izhikevich, Reeke, & Edelman, 2006).

Yet while modularity and integration are fundamental components of the organization of the human brain, various other essential properties of the brain have shaped its organization by implementing significant constraints. Chief among these is the evolutionary context. The human brain comprises a collection of specialized circuits that were selected for over time on the basis of environmental pressures and the social landscape, and they endow us with an unparalleled array of abilities and underlie our unique capacities. It is impossible to fully understand the intricate organization of the brain without addressing the forces that fashioned it. Furthermore, complex interactions between the splendid product of evolution—the human genome—and the environment that each individual encounters uniquely sculpt these circuits during development and throughout life. Patterns of activity in these circuits manifest as cognitive function and conscious experience, but fully understanding this activity requires characterizing the properties of neurons and their innumerable molecular constituents. Neuronal activity also has a metabolic cost; in fact, neurons have exceedingly high metabolic requirements (Raichle & Mintun, 2006) and consume energy, in the form of glucose, at an unsustainable rate during the waking hours of the day (Tononi & Cirelli, 2003).

Thus the biological brain is made of delicate organic tissue that, at the molecular and cellular levels, requires overwhelmingly precise homeostatic regulation (Marder & Prinz, 2002) and that, at the neural circuit and systems levels, produces an expansive range of adaptive human behavior and experience. A unified theory of brain function must parlay these separate principles and constraints into a comprehensive understanding of how the brain accomplishes this impressive feat. So far, the field has not approached anything converging on such an ideal, though different combinations of the above facts have received consideration (Sporns, Tononi, & Edelman, 2000b).

As we alluded to above, in the early going, we did not have the proper tools or intuitions to sustain such a comprehensive effort, so the foundation of the field was built by reducing grand mysteries to tractable puzzles, mainly by associating specific structures with observed functions. We believe that the time is right to step back and examine the larger framework within which the terrifically diverse discoveries of cognitive neuroscience fit. Sometimes the biggest hurdle in tackling seemingly insuperable problems is articulating what a solution might resemble. We attempt such a feat here in hopes of generating discussion, for it is often in the act of critique and revision that the deepest of insights are born.

We suggest that a critical step in a comprehensive understanding of the human brain is acknowledging that it is a complex system and thus exposing it to novel analysis from a somewhat different perspective. Though formal definitions of complex systems vary, there is a consensus that complex systems are composed of many independent elements that interact and, in doing so, generate emergent properties that are greater than the mere sum of the individual components. Complex systems are self-organizing and often are capable of adaptation, interacting with and changing on the basis of the environment (Amaral & Ottino, 2004). There is room for great variance in both the characteristics and number of building blocks of the system and the nature of their interactions, which may be dynamic or static, local or global. The claim that the brain is a complex system should not be controversial and has, in fact, been asserted by others (e.g. Sporns, Tononi, & Edelman, 2000a, 2006b; Jirsa & McIntosh, 2007). Given this characterization, it might be

informative to examine theoretical approaches to other complex systems. These discussions could provide important clues related to how to think about neural function. We turn now to a few complex systems that have received recent compelling theoretical treatment.

Other complex systems

By definition, all complex systems are composed of interconnected subunits, so at first glance, it might appear to be trivial to draw parallels between the brain and these systems. However, it is prudent, and quite useful, to dig a bit deeper and to examine specific complex systems that have architectural or functional properties similar to those of the human brain. For example, a Boeing 747 is a complex system made up of many interconnected parts, but to a large extent, it is unable to flexibly adapt to environmental demands and therefore might not provide much insight into the function of the brain (Amaral & Ottino, 2004). We examine complex systems that share attributes with the organization and function of the human brain beyond the obvious features dictated by the general criteria of complex systems, beginning with other examples of complex systems in nature.

A common feature of natural complex systems that is also observed in the brain is the dramatic effect that minor alterations in initial conditions can exert on global systems. Edward Lorenz (1963), a pioneer of meteorology and chaos theory, showed that large-scale dynamics in weather were highly sensitive to differences in initial conditions. This became known as the butterfly effect. Imagine, for example, that you were to place a leaf in a stream of rocky and turbulent water. Where you place the leaf, the angle at which you place it, and perhaps dozens of other incalculable variables will determine whether the leaf is carried downstream or ends up on one or other of the riverbanks. Thus small changes in initial conditions determine the final outcomes, which often cannot be predicted. One aspect of brain function is that large changes at the network level can be instituted by very small changes in initial conditions, such as individual spike rate behavior or the activity at the columnar level. While we know that the elemental processing unit of the brain is the neuron, groups of which self-organize into neuronal assemblies without any assistance from a central controller, we do not know how the organizational and functional structure of these assemblies produce observed emergent properties. Several studies and methods highlight the dynamic and complex nature of large-scale neural networks.

One recent study by Izhikevich and Edelman (2008) used diffusion tensor imaging (DTI) data to provide large-scale anatomical constraints on a model of brain function that reproduced spiking behavior of individual neurons that were capable of learning and organizing into functional assemblies. Interestingly, although the model's cortical microcircuitry was uniform, different frequency bands emerged at distinct areas of cortex. The differences in white matter connectivity provided by the DTI data likely accounted for the self-organizing behavior of the model. Additionally, they found that changing the spiking behavior of even a single neuron in the model changed the overall appearance of the large-scale network behavior. While weather systems and brain dynamics are linked at the level of being complex systems, it is clear that brain dynamics comprise subsystems coupled by functionally meaningful information (Jantzen & Kelso, 2007). A clearer understanding of human cognition will come from a greater understanding of how the individual components of the brain exchange information and work together to give rise to large-scale systems. Studying the thalamocortical system is proving to be fruitful in this endeavor (Tononi, 2004).

Other complex systems, such as ad auctions on search engines, share the distributed, decentralized architecture that we argue is a hallmark of brain organization. Consider the challenge faced by designers of a search engine such as Google, who are interested in using search inquiries to guide advertisement placement among search results. The sheer quantity of possible keywords negates the possibility of effectively setting prices for each keyword and ad position. A more viable solution is to use an auction, in which buyers largely set the price; however, traditional auctions feature a centralized auctioneer who ultimately determines the buyer and orchestrates the transaction. Such a structure is impractical—indeed, impossible—to implement when one considers the hugely diverse spectrum of potential advertising keywords and advertisers. Imagine the queue that would build up in such an instance; search engines would be rendered useless. Instead, innumerable simultaneous auctions occur in which Google, advertisers, and consumers engage in a common interaction despite unique interests. The beauty of the design of ad auctions is that the selfish motives of each party are harnessed and ultimately facilitate the most productive interaction possible (Varian, 2007). Advertisers place bids for target keywords, and each time a search occurs, the ad slots are filled on the basis of the highest bidder, who is charged only slightly more than the second highest bidder. An entire economy emerges, completely devoid of a centralized auctioneer.

The structure of ad auctions mirrors in many ways the very platform that is responsible for their existence, namely, the Internet. The Internet also consists of a distributed architecture in which independent processors, called autonomous systems, share processing demands while operating for varying purposes. The interdomain routing system connects these smaller systems, forming a sprawling, self-organizing network that lacks a "trusted center" (Feigenbaum, Papadimitriou, Sami, & Shenker, 2005). Information is sent in the form of packets, which often must traverse many

autonomous systems before reaching the system that is intended to receive them. Carrying this traffic is a burden to autonomous systems, but it is necessary for the function of the Internet, since there is no central system that connects to all autonomous systems and distributes information. Autonomous systems make computations that dictate routing patterns on the basis of information they receive from neighboring systems, leaving open the possibility that systems could lie about the costs they are incurring at any given moment to reduce the volume of packages they are sent. Lying would reduce efficiency of the overall network and thus be detrimental. Understanding how the Internet emerges from this mess of distributed computation and potentially devious and selfish local systems is an enormous challenge.

Complex systems such as ad auctions and the Internet warrant a novel theoretical approach, for they involve both independent, selfish agents (typically understood by the keen analysis of game theory) and complex computation (typically contained within the field of computer science). Moreover, in both cases, the fundamental structure precludes a central authority and requires distributed processing. An emerging theoretical field, distributed algorithmic mechanism design (DAMD) (Feigenbaum, Papadimitriou, & Shenker, 2001; Feigenbaum et al., 2005), represents a marriage of economic and computational concerns within a decentralized, distributed framework. DAMD is a valuable exemplar because it addresses three critical questions one must ask about a distributed complex system:

1. What drives self-organization of a complex system into such an architecture?
2. What are the characteristics of the nodes?
3. What are the nature of and constraints upon the interactions of the nodes?

In regard to the first question, DAMD identifies three major forces that compel a system to be distributed without a trusted center: trust, scalability, and reliability. Agents might not be willing to trust a centralized entity. Furthermore, it might be computationally infeasible for a center to exist, and in many cases, it is exceedingly risky to have a large system rely upon a single central node (Feigenbaum, Schapira, & Shenker, 2007). Addressing the attributes of different nodes, DAMD identifies four varieties: obedient nodes, faulty nodes, strategic (or selfish) nodes, and adversarial nodes (Feigenbaum & Shenker, 2002). Equations or models can be used to provide details or quantitative predictions about each node category. These nodes interact through a network, so a complete description of the system must also characterize these interactions. Complexity in this network is measured by the number of packages sent, the total number of packages sent over a single link, the maximum message size, the local computational burden on each node, and the storage required at nodes. For the network to function properly, each of these measures must be within a feasible range (Feigenbaum et al., 2007). Thus by determining the theoretical structure of complex systems such as the Internet or ad auctions, DAMD provides a general framework in which specific equations and measurements can be integrated to explain a given complex system.

The tenets of DAMD justify the distributed organization of the human brain promoted in this chapter and may provide a flavor of what a theory of brain function might resemble. First and foremost, the lack of a central trusted center resonates with the concept of a homunculus-less brain. The considerations of scalability and reliability are particularly relevant when we consider the brain, especially in an evolutionary context. A hallmark of the human brain is an increase in intrahemispheric white matter (Rilling & Insel, 1999), which indicates pressure toward specialization of independent processors (Striedter, 2005). As the brain accumulated specializations, the feasibility of a neural region listening in on and orchestrating activity in these diverse processors rapidly decreased. Furthermore, the distributed design is extremely reliable and thus adaptive. Focal neurological damage may incapacitate one or multiple modules, but in most cases, this does not profoundly disrupt conscious experience or the individual's conception of self (Tononi, Sporns, & Edelman, 1999). An extreme example is the split-brain patient, whose left hemisphere reports little disruption to its conscious world following callosal transection.

Though it might be possible to stretch this analogy further, it is important to emphasize that the purpose of this exercise is to derive useful ideas from the study of other complex systems rather than attempting to map neural function on a single existing theory. For instance, neural modules are not likely selfish in an economic sense, and they may be less predictable than obedient nodes, since they are sculpted by both genetic blueprints and interactions with the environment. However, the purpose of this brief tour through Boeings, streams, and the Internet was to expand the scope through which we peer at the brain, in hopes of gaining a better understanding of what a theory of global brain function might resemble. A comprehensive model should set out to do something similar to DAMD, explaining the origins and reasons for the organization of the system, explaining the properties of the nodes, and explaining their interactions. Note that a description of brain function that addresses this problem set will require contributions from evolutionary psychology, cognitive neuroscience, neurobiology, computational neuroscience, and many other fields. This should not be surprising. Assembling the brute force of these varied approaches and integrating their various perspectives in a unified theory that treats the brain as a complex system will surely be an exciting and fruitful venture.

Conclusion and future directions

New methodologies are providing an unprecedented opportunity to study the organization and function of the human brain and are giving rise to exciting results. Human neuroimaging studies have revealed that virtually every perceptual or cognitive task, whether it be object recognition, memory encoding and retrieval, reading, working memory, attentional processing, motor planning or awareness, is the result of activity within large-scale and distributed brain networks (McIntosh, 1999, 2000; Sporns & Tononi, 2007). Neuroimaging methodologies are not without their flaws, however. Recording behavioral measures and correlating them with "activation blobs" from functional imaging data leave enormous gaps in the larger picture. Activation blobs from these studies may be easily confused with spatially localized modules or nodes; however, they are likely to be part of a more distributed functional network. The spatial scale of functional magnetic resonance imaging (fMRI) studies has created an outflow of abbreviations designated, perhaps unintentionally, to be discrete modules, but the neuroimaging field has delivered little indication as to how these areas work together.

It is critical that we do not allow our methodologies to constrain our view of how the brain works. We believe that thinking of the brain as a complex system will help to circumvent perspectives that might otherwise limit our endeavors. To advance our understanding of its interconnected parts, we will need to view brain function as a series of dissociable networks, each communicating upward and downward with subsystems. Considering only the localization of certain functions is equivalent to focusing entirely on the properties of nodes in systems such as the DAMD architecture described above. It is critical that we look beyond functional localization to connectivity and the interaction between the two. Recall that the merit of approaches such as DAMD is the additional emphasis on the interactions between nodes, which ultimately give rise to emergent phenomena.

The importance of connectivity in the brain cannot be overlooked or underestimated. Three separate meanings exist in neuroimaging for the term *connectivity*: anatomical, functional, and effective. The term *anatomical connectivity* refers to the physical or structural connections that link neurons within a given network. Anatomical connections range in scale from local circuits (the minicolumn) to bidirectional white matter pathways, such as those linking frontal and parietal cortices (Schmahmann & Pandya, 2006). Methods such as diffusion-weighted imaging and tractography have allowed us, for the first time, to peer in vivo into the anatomical connections of the human brain (Le Bihan, 2003). Functional connectivity (Friston, Frith, Liddle, & Frackowiak, 1993) is a statistical measure of the correlation or covariance of activity in spatially distinct brain areas. This distributed activity can be quantified by using multiunit electrode recordings, EEG data, and even fMRI signals. However, measures of functional connectivity make no reference to the causal effect of one area over another. Finally, effective connectivity describes the causal effects one brain area has on another (Büchel & Friston, 2000). Integrating various neuroimaging analyses, to provide a more thorough account of the anatomical basis and functional significance of the interactions between distinct brain areas, will shed light on how the brain produces the enormous range of adaptive human behavior (Sporns, Tononi, & Edelman, 2002).

The need to integrate anatomical, functional, and effective connectivity challenges the field of cognitive neuroscience to think beyond localization of function and more toward a perspective that appreciates complex interactions. It is becoming increasingly clear that the brain operates on a large-scale network level (Sporns, Tononi, & Kotter, 2005; Hagmann et al., 2008). Recent work into what has become known as the brain's default-state network highlights the importance of approaching the brain as a complex system and emphasizing the interaction of connectivity and function. The default network is a system of highly interconnected brain areas that has consistently demonstrated deactivation when goal-directed behavior is needed (Buckner, Andrews-Hanna, & Schacter, 2008). Thus on the large scale, the brain seems to have at least two anticorrelated networks. One network, consisting of more lateral frontoparietal circuits, seems to be engaged during attention-demanding cognitive tasks. The other, the default network, is located in more medial regions and is involved in non-goal-oriented processes (Fox et al., 2005). Importantly, when the lateral circuits are not recruited by environmental demands, the default network becomes active, perhaps reflecting the functions of the idle mind. This illustrates that the brain possesses some kind of "circuit breaker" that switches and allocates substantial metabolic and neuronal resources in a context-dependent fashion. Efforts to understand how this is accomplished are underway and synthesize theoretical explanations for the origins of this architecture, characterizations of the critical cortical nodes, and a detailed understanding of how the three connectivity concepts noted above produce the emergent functionality (see chapter 73 in this volume).

The investigation of the default network represents a notable example of how thinking of the brain as a complex system produces tangible results. Ultimately, a great number of lower-level variables contribute to such large-scale networks at the level of neurons or neural circuits and their respective connections, and how these lower-level variables enable the emergent properties of the large scale is the question at hand. We can be certain that the sheer complexity of the human brain and the innumerable emergent states that it produces are directly tied to the vast

number of lower-level variables and interactions of which it consists. This is true of other complex systems as well. The Internet would be less impressive if it contained very few independent processors dedicated to only certain kinds of information. Imagine if the modern Internet were only Expedia.com.

No single theory pertaining to the analogous complex systems described above or any other complex system will encompass the findings of cognitive neuroscience, nor will it provide a comprehensive view of how the brain enables the mind. Instead, this burden falls on the shoulders of the researchers at the forefronts of neuroscience. And while the brilliant research, cutting-edge methodologies, and new insights into brain function contained in this volume breed tremendous optimism and excitement, there is still work to be done. Sometimes the thrill of progress dulls awareness of the big problems. As we move forward, it will be paramount not to get lost in details at the expense of a more comprehensive perspective on brain function. Furthermore, we need to acknowledge that the solution we seek should be based not on what was or what we think should be, but on what actually is. This will require communication across many fields, and these many contributions will need to be fit within a single framework that allows us to extract and distill significant ideas. We believe that thinking of the brain as a complex system can redistribute the collective attention of diverse fields in such a way that they will be primed to make inroads into some of the most enduring and tantalizing mysteries in all of science.

REFERENCES

Amaral, L. A. N., & Ottino, J. M. (2004). Complex networks: Augmenting the framework for the study of complex systems. *Eur. Phys. J. [B]*, *38*(2), 147–162.

Baynes, K., Eliassen, J. C., Lutsep, H. L., & Gazzaniga, M. S. (1998). Modular organization of cognitive systems masked by interhemispheric integration. *Science*, *280*(5365), 902–905.

Büchel, C., & Friston, K. (2000). Assessing interactions among neuronal systems using functional neuroimaging. *Neural Net.*, *13*(8–9), 871–882.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Ann. NY Acad. Sci.*, *1124*, 1–38.

Cooney, J. W., & Gazzaniga, M. S. (2003). Neurological disorders and the structure of human consciousness. *Trends Cogn. Sci.*, *7*(4), 161–165.

Crick, F., & Koch, C. (2000). The unconscious homunculus. In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions* (pp. 103–110). Cambridge, MA: MIT Press.

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nat. Neurosci.*, *6*(2), 119–126.

Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proc. Natl. Acad. Sci. USA*, *104*(26), 11073–11078.

Dux, P. E., Ivanoff, J., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved FMRI. *Neuron*, *52*(6), 1109–1120.

Feigenbaum, J., Papadimitriou, C., Sami, R., & Shenker, S. (2005). A BGP-based mechanism for lowest-cost routing. *Distrib. Comput.*, *18*(1), 61–72.

Feigenbaum, J., Papadimitriou, C. H., & Shenker, S. (2001). Sharing the cost of multicast transmissions. *J. Comput. Syst. Sci.*, *63*(1), 21–41.

Feigenbaum, J., Schapira, M., & Shenker, S. (2007). Distributed algorithmic mechanism design. In N. Nisan, T. Roughgarden, E. Tardos & V. V. Vazirani (Eds.), *Algorithmic game theory* (pp. 363–384). Cambridge, UK: Cambridge University Press.

Feigenbaum, J., & Shenker, S. (2002). *Distributed algorithmic mechanism design: Recent results and future directions*. Paper presented at the 6th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications, Atlanta.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. USA*, *102*(27), 9673–9678.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1993). Functional connectivity: The principal-component analysis of large (PET) data sets. *J Cereb. Blood Flow Metab.*, *13*(1), 5–14.

Gazzaniga, M. S. (1989). Organization of the human brain. *Science*, *245*(4921), 947–952.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.*, *6*(7), e159.

Holtzman, J. D., & Gazzaniga, M. S. (1982). Dual task interactions due exclusively to limits in processing resources. *Science*, *218*(4579), 1325–1327.

Izhikevich, E. M., & Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proc. Natl. Acad. Sci. USA*, *105*(9), 3593–3598.

Jantzen, K. J., & Kelso, J. A. (2007). Neural coordination dynamics of human sensorimotor behavior: A review. In V. K. Jirsa & A. R. McIntosh (Eds.), *Handbook of brain connectivity* (pp. 421–462). New York: Springer.

Jirsa, V. K., & McIntosh, A. R. (2007). *Handbook of brain connectivity*. New York: Springer.

Le Bihan, D. (2003). Looking into the functional architecture of the brain with diffusion MRI. *Nat. Rev. Neurosci.*, *4*(6), 469–480.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. USA*, *99*(14), 9596–9601.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.*, *20*, 130–141.

Marder, E., & Prinz, A. A. (2002). Modeling stability in neuron and network function: The role of activity in homeostasis. *Bioessays*, *24*(12), 1145–1154.

McIntosh, A. R. (1999). Mapping cognition to the brain through neural interactions. *Memory*, *7*(5–6), 523–548.

McIntosh, A. R. (2000). Towards a network theory of cognition. *Neural Net.*, *13*(8–9), 861–870.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends Cogn. Sci.*, *7*(3), 141–144.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Raichle, M. E., & Mintun, M. A. (2006). Brain work and brain imaging. *Annu. Rev. Neurosci.*, *29*, 449–476.

Rilling, J. K., & Insel, T. R. (1999). Differential expansion of neural projection systems in primate brain evolution. *NeuroReport*, *10*(7), 1453–1459.

Schmahmann, J. D., & Pandya, D. N. (2006). *Fiber pathways of the brain*. New York: Oxford University Press.

Seth, A. K., Izhikevich, E., Reeke, G. N., & Edelman, G. M. (2006). Theories and measures of consciousness: An extended framework. *Proc. Natl. Acad. Sci. USA*, *103*(28), 10799–10804.

Sporns, O., & Tononi, G. (2007). Structural determinants of functional brain dynamics. In V. K. Jirsa & A. R. McIntosh (Eds.), *Handbook of brain connectivity* (pp. 117–148). New York: Springer.

Sporns, O., Tononi, G., & Edelman, G. M. (2000a). Connectivity and complexity: The relationship between neuroanatomy and brain dynamics. *Neural Net.*, *13*(8–9), 909–922.

Sporns, O., Tononi, G., & Edelman, G. M. (2000b). Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cereb. Cortex*, *10*(2), 127–141.

Sporns, O., Tononi, G., & Edelman, G. M. (2002). Theoretical neuroanatomy and the connectivity of the cerebral cortex. *Behav. Brain Res.*, *135*(1–2), 69–74.

Sporns, O., Tononi, G., & Kotter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Comput. Biol.*, *1*(4), e42.

Striedter, G. F. (2005). *Principles of brain evolution*. Sunderland, MA: Sinauer Associates.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.*, *5*, 42.

Tononi, G., & Cirelli, C. (2003). Sleep and synaptic homeostasis: A hypothesis. *Brain Res. Bull.*, *62*(2), 143–150.

Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. USA*, *91*(11), 5033–5037.

Tononi, G., Sporns, O., & Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proc. Natl. Acad. Sci. USA*, *96*(6), 3257–3262.

Varian, H. R. (2007). Position auctions. *Int. J. Ind. Organ.*, *25*(6), 1163–1178.

Volpe, B. T., Ledoux, J. E., & Gazzaniga, M. S. (1979). Information processing of visual stimuli in an "extinguished" field. *Nature*, *282*(5740), 722–724.

Wardak, C., Ibos, G., Duhamel, J. R., & Olivier, E. (2006). Contribution of the monkey frontal eye field to covert visual attention. *J. Neurosci.*, *26*(16), 4228–4235.

89 The Landscape of Cognitive Neuroscience: Challenges, Rewards, and New Perspectives

ELISSA M. AMINOFF, DANIELA BALSLEV, PAOLA BORRONI, RONALD E. BRYAN, ELIZABETH F. CHUA, JASMIN CLOUTIER, EMILY S. CROSS, TRAFTON DREW, CHADD M. FUNK, RICARDO GIL-DA-COSTA, SCOTT A. GUERIN, JULIE L. HALL, KERRY E. JORDAN, AYELET N. LANDAU, ISTVAN MOLNAR-SZAKACS, LEILA MONTASER-KOUHSARI, JONAS K. OLOFSSON, SUSANNE QUADFLIEG, LEAH H. SOMERVILLE, JOCELYN L. SY, LUCINA Q. UDDIN, AND MAKIKO YAMADA

The chapters in the volume you are holding reflect an ever-evolving understanding of how the brain—from genes to proteins, from cells to systems, and building blocks in between—generates behavior. Before publication of this book, 22 fellows of the Summer Institute in Cognitive Neuroscience attended a three-week meeting centered on the topics within this text and were challenged by Michael Gazzaniga to write a commentary reflecting current issues in the field. It is not an easy task to integrate a variety of conceptual and methodological approaches and produce an insightful commentary while escaping oversimplification of the issues. Nonetheless, the enterprise proved both exciting and worthwhile, and we hope that it illuminates and unifies some seemingly disparate concepts in a new light. Consistent with the nature of the cognitive neurosciences, we are a heterogeneous, multidisciplinary group. It is from this perspective, by tying together themes that transcend traditional research topics, that we present here our own reflections as we look toward the future of the field.

First, we will discuss recent methodological developments that have allowed for greater integration across multiple levels of analysis. These methods have expanded the range of research questions that can be asked and in some cases have led to new theoretical approaches. In turn, novel theoretical models have generated profound shifts in research foci. Cognitive neuroscientists are now exploring topics that were previously considered impossible or implausible for scientific investigation (e.g., social cognitive neuroscience) and are also revisiting old themes with a new mindset (e.g., the contribution of nature versus nurture). To conclude, we reflect on how this "new" cognitive neuroscience is beginning to influence everyday life, including public policy, education, and healthcare.

Summer Institute for Cognitive Neuroscience 2008, Lake Tahoe, California. All coauthors contributed equally to this chapter

Methods

The evolution of cognitive neuroscience has been driven largely by the development of increasingly sophisticated experimental methods. These methods and novel techniques enable us to address hypotheses that were previously unimaginable, such as establishing a causal relationship between patterns of neural activity, cognitive processes, and complex behavior. Methods that are currently under development are allowing us to better estimate the spatiotemporal structure of neural data, move from correlation-based methods to experimental manipulation of brain activity, and gain a deeper understanding of the importance and instructive value of individual differences.

One significant outcome of recent technological advances is the ability to quantitatively measure brain structure (e.g., diffusion tensor imaging and virus tracing technology) and function (e.g., high-resolution functional magnetic resonance imaging (fMRI) and single-unit recordings in humans) at a higher spatial and/or temporal resolution than ever before. The enhanced precision of these measurements is accompanied by increasingly sophisticated data analysis techniques, computational models, and theoretical interpretations. For example, novel multivariate and pattern classification approaches to fMRI data have revealed a spatial structure of hemodynamic activity beyond what is evident from traditional univariate approaches (Carlson, Schrater, & He, 2003; Cox & Savoy, 2003).

One challenge that remains is to bring these methods to bear on the long-standing issue of modular versus distributed processing networks in the brain. An increasing number of

multivariate analysis techniques enable us to describe brain activity on a network level, including independent and principal component analyses, dynamic causal modeling, and multivoxel pattern analysis. Such methods allow us to test existing theoretical ideas but also reveal new issues that were previously not accessible to scientific measurement. For example, pattern classification techniques have reframed the issue of object processing in the ventral temporal cortex from *where* an object is represented to *how* patterns of brain activity represent information (Norman, Polyn, Detre, Haxby, 2006). We anticipate further methodological advances, as increased resolution reveals new scales of organization of the human brain.

A key objective of any scientific investigation is to uncover causal structure. Continuing advances in functional brain imaging enable neuroscientists to infer functional-anatomical correlates from patterns of blood flow during task performance, but these methods are limited to uncovering correlational links, not causal ones. Historically, the ability to draw conclusions regarding the necessity of a brain structure to perform a particular cognitive operation has been limited to lesion studies in animals and the rare occurrence of focal lesions in humans due to brain insults or surgery. In recent years, new methodological techniques such as transcranial magnetic stimulation (TMS) have enabled the noninvasive manipulation of cortical activity in healthy individuals (chapter 9 in this volume). Additionally, in neurosurgical settings, researchers are increasingly using electrical microstimulation to directly activate a particular population of neurons and study its behavioral outcomes, as well as changes in the neurophysiology locally (around the electrode) and in other parts of the brain that are connected to the stimulated region, (e.g., Kawasaki & Sheinberg, 2008; Houweling & Brecht, 2008). On a different scale, gene knockout manipulations in mice reveal remarkable insights into the genetic basis of cognition and behavior (e.g., Hung et al., 2008; Gunnerson et al., 2007). For example, Hung and colleagues found enhanced performance in spatial learning but impaired long-term retention in Shank1 knockout mice, suggesting an important and complex role for the Shank family of proteins in normal cognitive development. Another approach using neuropharmacological methods has led to exciting new data that point to a link between serotonin levels in the brain and interpersonal trust among human participants (Crockett, Clark, Tabibnia, Lieberman, & Robbins, 2008). As we continue to work toward the refinement of this diverse toolbox of experimental approaches, we are hopeful that they will help to further elucidate the causal links between brain function, structure, and behavior.

An emerging trend in cognitive neuroscience methodology is the study of how individual differences contribute to our understanding of the human brain. Psychologists and neuroscientists have conventionally relied on aggregating data across a sample of subjects to draw conclusions about how the "average" or prototypical brain might function. In reality, individual differences abound in the dynamic interplay between environment, genes, behavior, anatomy, and topographic patterns of brain activity. These differences offer a vast and scarcely explored source of clues about the nature of the mind and brain. To this end, new methods in correlating environmental and genetic variation with neural data offer exciting new possibilities in understanding how genes and the environment interact to produce a unique brain (see chapters 50 and 64 in this volume). Accounting for interindividual variance not only may give us a more ecologically valid way of looking at cognition, but also may provide important insights into more general, overarching processes that underlie different types of cognition across individuals and perhaps even species. Perplexing as it might seem, understanding universal neural architecture that creates the human condition could rely on first recognizing the complex processes that make each of us unique.

Theory

As reflected in the evolution of the topics addressed in *The Cognitive Neurosciences* volumes that have been published over the past 20 years, there have been major advances in how we conceptualize mind, brain, and body interactions. New methods described above have also led to groundbreaking findings related to genetics and synaptic physiology and to the functions and connectivity underlying complex neural networks. One way to better understand the complexity of the brain has been to further integrate levels of analyses, research perspectives, and conceptual approaches.

For example, genes and experience are no longer considered to be polar opposite factors of influence but are instead viewed as dynamically interacting. The emergent view is that genes serve as scaffolding for the ways in which experience can change brain organization and subsequent behavior. Our innate genetic blueprint determines the development of the human brain continuously throughout the life span at various neural levels (e.g., directing cellular organization, pruning, myelination, molecular structures, cortical maturation, and connectivity), laying the biological foundation for behavioral functionality. For example, various genes have been implicated in anomalous maturation of neural systems and developmental disorders (e.g., FOXP2, among other genes in developmental dyslexia) (Ramus, 2006) and have been linked to the presence of specific psychiatric conditions (e.g., abnormal pruning in schizophrenia and autism) (McGlashan & Hoffman, 2000; Boylan, Blue, & Hohmann, 2007). As our knowledge of gene-environment interactions advances, it becomes ever more evident that while specific genotypes are associated with the phenotypic expression of

certain behaviors, environmental factors are also required for the expression of particular genotypes.

An increasingly sophisticated understanding of these gene-environment interactions is one realm in which the concept of dynamic interplays has borne fruitful insights. One can also observe this dynamic interplay at the level of the brain, which is an ever-changing functional system regulated by adaptive feedback mechanisms. Recent experiments have explored the boundaries of this functional flexibility by looking at primary sensory areas, for which the functional architecture was previously thought to be hard-wired. Specifically, Sur and colleagues have shown that if visual input in the ferret is experimentally rewired during development to auditory cortex, there is an emergence of orientation tuning columns in those auditory areas (Majewska & Sur, 2006), an elegant demonstration that the experiential input to primary sensory areas can dramatically alter functional organization.

However, although the input to a brain area has a powerful ability to remodel cortical structure and function, it must act in concert with the scaffolding laid down by genes. Thus while the rewired auditory cortex shares a similar structure with V1, its orientation map was found not to be as regular and precise, showing that while the brain has the capacity for remarkable plasticity, future research will have to delineate the boundaries of this adaptive function.

At a higher level of cognitive operations, adaptive feedback mechanisms have profound influences on brain-behavior relations. This flexibility is essential to keep up with our ever-changing social environments through the adaptive implementation of appropriate behavioral responses. Accordingly, current research emphasizes the role of top-down control over the selection of information relevant to our behavioral goals. These top-down factors are now believed to modulate cognitive processes at many levels. The influence of top-down control can be illustrated by the activity of sensory neurons, which are modulated by task goals. For example, the neural response of early visual cortices has been shown to be upregulated prior to the onset of attentional cues (Hopfinger, Buonocore, & Mangun, 2000; Giesbrecht, Woldorff, Song, & Mangun, 2003). Top-down modulation has also been shown to prepare sensory areas for upcoming stimuli by filtering out, or suppressing, the processing of unnecessary information and enhancing the processing of relevant features as far downstream as the lateral geniculate nucleus within the visual system (O'Connor, Fukui, Pinsk, & Kastner, 2002). This act of filtering information not only has important implications in the sensory processing of a stimulus, but also can have a strong impact on how we remember a past event. For example, if an emotional event is depicted within a broader spatial context, our attention is focused on the emotional content, leading recognition memory for the background information to suffer (Kensinger, Piguet, Krendl, & Corkin, 2005).

Top-down and attentional influences have been shown to affect not only systemwide neural activity, but also the precise timing of neural firing patterns within local populations of neurons. For example, increased attention enhances local gamma band synchrony or temporally coincident firing within a population of neurons in the motor system, which, in turn, upregulates the signal output from that region, thereby increasing the speed of behavioral responding (Fries, Womelsdorf, Oostenveld, & Desimone, 2008). These findings demonstrating the flexibility of brain and behavior relationships are a product of integrating different levels of analysis, from genetic expression to neurophysiology to functional organization to manifestations in behavior.

Given the recent trend toward integration of various levels of analysis, there is an emergent move to study neural activity in networks—both local and global—as well as systems across the whole brain. Let's consider an example: Brain regions involved in processing different features of an object (e.g., color, shape, motion trajectories) are reactivated when retrieving visual detail about the object from memory (Martin, 2007). This demonstrates that information about objects is represented via a distributed network in the brain, and various processes related to using such information rely on this distributed network. To fully account for such effects, cognitive processes and their underlying neuronal mechanisms must be analyzed on a global scale. Specifically, traditional neuroimaging studies have focused on the "subtraction paradigm," in which functional localization is inferred from the brain's response to a particular cognitive task. A shift from localization to understanding large-scale network interactions parallels the emergence of the idea that one can achieve a greater understanding of neural processes by examining dynamic interactions between brain regions. Increasing interest in this network approach is reflected in the growing popularity of integrating multiple neuroimaging techniques in order to understand the structural and spatial (fMRI and diffusion tensor imaging), functional (fMRI, TMS, lesion, and event-related potential (ERP)), and temporal (electroencephalography and ERP) aspects of neural processes.

Another new approach is to study large-scale networks by examining spontaneous, task-independent fluctuations in blood oxygen level dependent (BOLD) signal that can reveal the intrinsic functional architecture of the brain. The first studies by Raichle and colleagues (2001) using positron emission tomography and later fMRI highlighted the fact that spontaneous neuronal activity accounts for the majority of the brain's energy metabolism. In the last several years, the number of studies using the recently developed method of resting-state functional connectivity, which examines temporal correlations between discrete brain regions, has increased

exponentially. While the exact nature and function of so-called resting-state networks and their relationship with anatomical connectivity remain to be elucidated, the field has already progressed to the point at which it is clear that resting-state functional connectivity measures show meaningful correlations with behavioral measures and can be used to understand aberrant cortical connectivity in different clinical populations (Castellanos et al., 2008). Furthermore, this focus on the resting state not only enables analysis of multiple neural networks, but also could have important implications cutting across various fields of cognition. Thus we see the recent emphasis on examination of spontaneous brain activity as an important theoretical advancement with the potential to open new avenues of research in the years to come.

Integration of different levels of analyses has also borne fruit in the fledgling field of social cognitive and affective neuroscience. These efforts have been in part facilitated by the convergence of perspectives (e.g., neuroscience, social psychology, economics) towards understanding higher-order functions of the human mind. As humans, we constantly reflect on ourselves and regulate our behaviors, thoughts, and emotions. Building on work in many areas of neuroscience (e.g., the study of the neural circuitry associated with various emotional experiences), advances have been made in identifying the role of brain areas involved with self-reflection and self-regulation, broadly defined. For example, recent work has started to reveal networks of brain areas that are recruited when we reflect about ourselves in the past, the present, and the future (Arzy, Molnar-Szakacs, & Blanke, 2008; Schacter, Addis, & Buckner, 2007). Another essential aspect of the human mind is our ability to navigate our rich and complex social environment. Indeed, networks of brains areas (i.e., medial prefrontal cortex, superior temporal sulcus, temporoparietal junction, fusiform, intraparietal sulcus, and amygdala) are now believed to be involved in the perception and understanding of others (i.e., their faces, actions, intention, and mental states) (Molnar-Szakacs & Arzy, 2009; Gobbini & Haxby, 2007; Mitchell, 2008). One conceptual breakthrough has been the idea that we use overlapping neural systems in representing our own knowledge, beliefs, intentions, and actions as we use to understand others. This "simulation" account finds empirical support in work by Rizzolatti and colleagues, who have shown that a special class of neurons located in frontal and parietal cortices in the macaque (e.g., "mirror neurons") respond to both executed and observed actions (Rizzolatti & Craighero 2004). Studies using neuroimaging techniques in humans have revealed a parallel frontoparietal mirror system that has been implicated in a variety of high-level cognitive and socioemotional processes that may rely, at least in part, on a simulation mechanism, including imitation and intention understanding (Iacoboni et al., 2005). These social cognitive and affective neuroscience approaches are already moving the field in the direction of bringing more ecological validity to the study of cognitive neuroscience, paralleling the social world in which we live.

Integrating the cognitive neurosciences

The common underlying question shared by the many lines of research collected in the chapters of this book can perhaps be summarized as "How does the human brain integrate multiple levels of neural responses—from molecules to neurons, to circuits—to produce adaptive behavior?" Immediately, a second-order question comes to mind: "How do *we* make sense of all the information that we have acquired about these systems in the past 100 years?" While reading the previous sections on the current theoretical and methodological advances in the field, three key themes seem to stand out with striking relevance: *time*, *context* and *integration*.

Although none of the themes is new per se, they reemerge in current research and shape the landscape of the field in new ways. The necessity for an integrated approach to the space-time dimensions has always been a fundamental framework in the study of cognitive neurosciences. Recently, in addition to the quest for localization of function, an increased emphasis is placed on describing the temporal unfolding of neural activity, such as elucidating neural networks sharing temporal synchrony.

As such, *time* can be seen as a defining element of change, a critical feature to both brain physiology and behavioral output. For example, change in molecular constitution and neural connectivity determines the acquisition of new functional properties during development, perception, learning, memory encoding and retrieval, attention, and every other cognitive activity that one might consider. From an ontogenetic perspective, the temporal unfolding of plasticity critically shapes the growing organism's cognitive machinery, from cellular mechanisms such as specialized neuronal growth and wiring to later developmental stages associated with higher cognitive functions such as language.

In a phylogenetic perspective, the timescale expands to include the historical record of progress toward understanding the biological bases of cognition. Thus the importance of the study of multiple biological models, with last common ancestors at different evolutionary times, has become increasingly recognized (for coverage of this topic, see Platek, Keenan, & Shackleford, 2006).

The consideration of *context* in current research emerges as an increasingly exciting dimension in cognitive neuroscience. It is now evident that the substrates of cognitive functions cannot be studied in isolation but that insight into their mechanisms and consequent outcomes can be gained only from the full contextual setting in which they develop and operate. From neuronal activity being strongly modulated

by the extracellular surround and glial support to the environmental context influencing the processing of a given stimulus, context plays a critical role in driving and explaining neural activity. One can also consider context more broadly, such as the evolutionary forces that have shaped human and nonhuman primate brains and behavior, driven by ecological and environmental factors. From cellular to systems to social neurosciences, there is no scarcity of examples demonstrating the importance of context.

Finally, the pages of this book consistently remind us of the overwhelming importance of integration. We believe that now is an excellent time for integration at all levels! Critical to the advancement of the field is integration in many forms: of methods, of biological models, and of conceptual perspectives, all reflected in the integration of researchers with very different scientific backgrounds investigating the complexity of the multifaceted discipline of cognitive neuroscience.

Clearly, the advancement of cognitive neuroscience strongly relies on the integration of knowledge across perspectives and domains. The attempt to overcome these limits is already evident in the current use of multiple techniques and interdisciplinary approaches, as well as in the increasing number of multidisciplinary teams and research collaborations. An integrative approach is fundamental to the achievement of our common goal: understanding the multitude of complex neural systems at the core of the causal relationship between our brain and our behavior.

Cognitive neuroscience and society

Over the past two decades, the field of cognitive neuroscience has developed at a frenetic pace. A new wave of research has moved past basic work aimed at understanding brain function toward examining the neural underpinnings of issues that are critically important to humanity. This is admittedly a lofty goal, but cognitive neuroscience has already made important contributions across several spheres of society. In conclusion, we highlight two specific cases in which the study of the brain clarifies and enhances our understanding and social well-being: uncovering the etiology and treatment mechanisms of psychiatric illness and informing behavioral interventions to enhance classroom learning for at-risk young children. While illuminating some societal issues, applied cognitive neuroscience research has complicated others, such as giving rise to legal controversies by fueling debates within the criminal justice system regarding culpability for our actions given certain brain insults. Finally, the surge in applied work has also placed neuroscience research squarely in the public eye. We will conclude with a discussion of the heightened responsibility required of scientists to communicate their findings to the general public in an informative and accurate fashion.

Cognitive neuroscience research is making important contributions in a number of clinical domains, notably in our understanding of the etiology of psychiatric disorders such as depression. For example, brain research has led to the development of animal models that provide a unique window into the cellular and molecular mechanisms of this disorder (Fuchs & Fliugge, 2006; McArthur & Borsini, 2006). In humans, neuroimaging techniques in combination with genetic analyses have assisted in characterizing both the structural and functional profiles associated with affective disorders. For instance, the serotonin transporter gene linked polymorphic region (5-HTTLPR) has been identified as a predictor of vulnerability for affective disorders and exaggerated amygdala response to emotional stimuli, as observed by using fMRI techniques (Munafo, Brown, & Hariri, 2008, Hariri et al., 2002). Patterns of fMRI activity can also predict the probability of remission from clinical depression and potentially guide treatment choices. Canli and colleagues (2005) found that greater amygdala activation to emotional faces predicts subsequent symptom reduction in depressed patients, identifying a subgroup of individuals who are predicted to have poorer chances of spontaneous alleviation of depressed symptoms. Furthermore, a number of experiments suggest that repetitive TMS to certain cortical regions might actually exert clinically significant antidepressant effects. Several studies from the laboratory of Alvaro Pascual-Leone (e.g., Stern, Tormos, Press, Pearlman, & Pascual-Leone, 2007) found that repetitive stimulation of the left or right dorsolateral prefrontal cortex was associated with symptom reduction in individuals with recurrent unipolar depression when compared with sham stimulation. These lines of research provide exciting evidence and hope that cognitive neuroscience will inform the prevention, diagnosis, and treatment of severe and disabling conditions of the brain and mind.

Advances in the fields of cognition and neuroscience also shed light on our understanding of brain processes such as learning and memory, which are critical to educational practices. Recent research topics include visual perception benefits accrued from playing complex video games (Green & Bavelier, 2003), focused attention on interventions for both children and parents, reading interventions during the development of language perception, and possible neural correlates of women's underperformance in math (e.g., Stevens and Neville, 2008; Varma, Mc Candliss, & Schwartz, 2008). Well suited to integrating the once disparate disciplines of neuroscience and education, cognitive neuroscience is even elucidating neural algorithms for specific scholastic subject areas. For example, in the domain of mathematics, this approach has begun to identify the brain bases of numerical thinking. Behavioral and neuroimaging results suggest that children and adults share a "number sense," that is, a system for representing approximate numerical magnitude

(Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Jordan & Brannon, 2006). It has recently been found that this neural system is compromised in the approximately 5% of children who have developmental dyscalculia, a specific mathematical learning disability (Price, Holloway, Vesterinen, Rasanen, & Ansari, 2007).

Such neuroscientific findings encourage very early interventions to target deficits, even before children enter elementary grades. One such recent behavioral intervention is to play linear number board games akin to Chutes and Ladders, which was shown to boost numerical magnitude understanding in preschoolers who were at risk for later falling behind in mathematics (Ramani & Siegler, 2008). This approach is one example of how cognitive neuroscience is helping in the field of education by translating neuroscientific findings into cognitive interventions that can effectively be implemented at an early age.

Another interesting implication of cognitive neuroscience is its intersection with the law (Gazzaniga, 2005). In a famous court case in 1982, John Hinckley, Jr., who had attempted to assassinate President Ronald Reagan, submitted CT scans indicating cortical atrophy in an attempt to gain a verdict of not guilty due to insanity. In a controversial decision, Hinckley's defense ultimately succeeded (*U.S. v. Hinckley* 1982). Since this historical case, evidence based in cognitive neuroscience, albeit remaining highly controversial, has played an increasingly important role in the criminal justice system. For example, fMRI-based lie detection tests have become a part of legal proceedings (Kozel et al., 2005; Gamer, Bauermannn, Stoeter, & Vossel, 2007), although empirical evidence for their reliability is still in question. While providing an illuminating example of the integration of cognitive neuroscience and the law, it is important to ensure that such progress does not compromise justice (Poldrack, 2008). Most important, the legal concept of responsibility should not be weakened by an endless flood of legal defenses attempting to reduce culpability on the basis of subtle neural defects (Gazzaniga, 2005).

Another concern centers on the issue of privacy (Tovino, 2007). Neuroimaging methods allow us to extract health-related information based on patterns of neural activity and brain structure. For instance, we currently have the ability to predict cognitive decline (Small et al., 2008) and risk for psychiatric illness (Phillips & Vieta, 2007) on the basis of the detection of biomarkers using neuroimaging. It is likely that the use of fMRI in the private sector will only increase in the coming years. Laws to regulate the use of private health-related information and to protect the privacy of individuals will have to be enacted.

These are only a few of the ethical challenges faced by those whose job it is to update societal institutions and policies with the knowledge gained from cognitive neuroscience. The newly emerging field of neuroethics should play a vital role in protecting society from the detrimental effects or imprecise applications of neuroscience and must thoroughly address the difficult ethical dilemmas that arise.

To encourage public dialogue and awareness of the philosophical, ethical, and practical implications of neuroscientific research, U.S. president George H. W. Bush declared the 1990s to be the Decade of the Brain. Whether it was this proactive step on the part of the American government or the rapid and prolific development of the field of cognitive neuroscience, public interest in the brain has exploded. In the entire decade between 1980 and 1990, a mere 100 books were published related to cognitive neuroscience; this number increased tenfold during the Decade of the Brain to more than 1000 books. Since the year 2000, this figure has multiplied to over 4000 volumes on cognitive neuroscience in print (as determined by a keyword search of "cognitive neuroscience" on www.amazon.com in June 2008). Many of these books were written for a lay audience and have become mass market paperbacks reaching a public that is fascinated by the brain—and rightfully so. Cognitive neuroscience has the potential to explain the neural processes underlying consciousness, free will, morality, sexuality, emotion—all of those seemingly ineffable qualities that set us humans apart from other animals. As a result, many new research findings have implications that reach far beyond the relatively small cognitive neuroscience research community. The unique position of cognitive neuroscience of being a rather public science cannot be taken lightly, as we have the potential to profoundly alter and reshape the perception of the human condition in the public eye.

The pressure to present complex empirical data and theories so that they are accessible to nonexperts has sometimes led to excessive reductionism in the popular media. Catchy headlines such as "Nose cells may help paralyzed to walk again," "This is your brain on politics," and "'God spot' researchers see the light in MRI study" attempt to generate interest in readers but provide inaccurate "just-so" explanations of brain function. Admittedly, it is difficult to balance the goal of providing straightforward answers to difficult questions with effective communication to a broad audience, even for those of us in the field. Consequently, we as researchers must work together with journalists, writers, and reporters in the media to recognize and embrace our responsibility to help experts, as well as nonexperts, understand and appreciate the pluralism of our data and of the human mind. Cognitive neuroscience is an evolving discipline, and that is what makes it so exciting, but it also means that insights are sometimes fragmentary, methods suffer from shortcomings, and theories are often conflicting. Accurately representing and communicating these complexities will allow the public to see the passion but also the tireless effort of cognitive neuroscientists. We should therefore be eager to invite

society on this exciting journey toward a deeper understanding of how the human brain creates a mind that has the potential to fall ill, to forget, or to lose sense but also to reason, to empathize, to dream, and to love.

REFERENCES

Arzy, S., Molnar-Szakacs, I., & Blanke, O. (2008) Remembering the future, predicting the past: a neuroimaging study of mental time travel. *J. Neurosci.*, *28*(25), 6502–6507.

Boylan, C. B., Blue, M. E., & Hohmann, C. F. (2007). Modeling early cortical serotonergic deficits in autism. *Behav. Brain Res.*, *176*(1), 94–108.

Canli, T., Cooney, R. E., Goldin, P., Shah, M., Sivers, H., Thomason, M. E., et al. (2005). Amygdala reactivity to emotional faces predicts improvement in major depression. *NeuroReport*, *16*(12), 1267–1270.

Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *J Cogn. Neurosci*, *15*(5), 704–717.

Castellanos, F. X., Margulies, D. S., Kelly, C., Uddin, L. Q., Ghaffari, M., Kirsh, A., et al. (2008). Cingulate-precuneus interactions: A new locus of dysfunction in adult attention deficit/hyperactivity disorder. *Biol. Psychiatry*, *63*(3), 332–337.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2, Pt 1), 261–270.

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., & Robbins, T. W. (2008). Serotonin modulates behavioral reactions to unfairness. *Science*, *320*(5884), 1739.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974.

Fries, P., Womelsdorf, T., Oostenveld, R., & Desimone, R. (2008). The effects of visual stimulation and selective visual attention on rhythmic neuronal syncronization in macaque area V4. *J. Neurosci.*, *28*(18), 4823–4835.

Fuchs, E., & Fliugge, G. (2006). Experimental animal models for the simulation of depression and anxiety. *Dialogues Clin. Neurosci.*, *8*(3), 323–333.

Gamer, M., Bauermann, T., Stoeter, P., & Vossel, G. (2007). Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information. *Hum. Brain Mapp.*, *28*(12), 1287–1301.

Gazzaniga, M. S. (2005). *The ethical brain*. New York: Dana Press.

Giesbrecht, B., Woldorff, M. G., Song, A. W., & Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *NeuroImage*, *19*, 496–512.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of visually familiar faces. *Neuropsychologia*, *45*(1), 32–41.

Green, C. S., & Bavelier, D. (2003). Action video games modify visual attention. *Nature*, *423*(6939), 534–537.

Gunnersen, J. M., Kim, M. H., Fuller, S. J., De Silva, M., Britto, J. M., Hammond, V. E., et al. (2007). Sez-6 proteins affect dendritic arborization patterns and excitability of cortical pyramidal neurons. *Neuron*, *56*(4), 621–639.

Hariri, A. R., Mattay, V. S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M. F., & Weinberger, D. R. (2002). Serotonin transporter genetic variation and the response of the human amygdala. *Science*, *297*(5580), 400–403.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat. Neurosci.*, *3*, 284–289.

Houweling, A. R., & Brecht, M. (2008). Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, *451*(7174), 65–68.

Hung, A. Y., Futai, K., Sala, C., Valtschanoff, J. G., Ryu, J., Woodworth, M. A., et al. (2008). Smaller dendritic spines, weaker synaptic transmission, but enhanced spatial learning in mice lacking Shank1. *J. Neurosci.*, *28*(7), 1697–1708.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *Publ. Library Sci. Biol.*, *3*(3), 529–535.

Jordan, K., & Brannon, E. (2006). The multisensory representation of number in infancy. *Proc. Natl. Acad. Sci.*, *103*, 3486–3490.

Kawasaki, K., & Sheinberg, D. L. (2008). Learning to recognize visual objects with microstimulation in inferior temporal cortex. *J. Neurophysiol.*, *100*(1), 197–211.

Kensinger, E. A., Piguet, O., Krendl, A. C., & Corkin, S. (2005). Memory for contextual details: Effects of emotion and aging. *Psychol. Aging*, *20*(2), 241–250.

Kozel, F., Johnson, K., Mu, Q., Grenesko, E., Laken, S., & George, M. (2005). Detecting deception using functional magnetic resonance imaging. *Biol. Psychiatry*, *58*(8), 605–613.

Majewska, A., & Sur, M. (2006). Plasticity and specificity of cortical processing networks. *Trends Neurosci.*, *29*, 323–329.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25–45.

McArthur, R., & Borsini, F. (2006). Animal models of depression in drug discovery: A historical perspective. *Pharmacol. Biochem. Behav.*, *84*(3), 436–452.

McGlashan, T. H., & Hoffman, R. E. (2000). Schizophrenia as a disorder of developmentally reduced synaptic connectivity. *Arch. Gen. Psychiatry*, *57*(7), 637–648.

Mitchell, J. P. (2008). Contributions of functional neuroimaging to the study of social cognition. *Curr. Dir. Psychol. Sci.*, *17*, 142–146.

Molnar-Szakcs, I., & Arzy, S. (2009). Searching for an integrated self-representation. *Commun. Integr. Biol.*, *2*(4), 1–3.

Munafo, M. R., Brown, S. M., & Hariri, A. R. (2008). Serotonin transporter (5-HTTLPR) genotype and amygdala activation: A meta-analysis. *Biol. Psychiatry*, *63*(9), 852–857.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.*, *10*(9), 424–430.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attentional modulates responses in the human lateral geniculate nucleus. *Nat. Neurosci.*, *5*(11), 1203–1209.

Phillips, M. L., & Vieta, E. (2007). Identifying functional neuroimaging biomarkers of bipolar disorder: Toward *DSM-V*. *Schizophr. Bull.*, *33*(4), 893–904.

Platek, S. M., Keenan, J. P., & Shackelford, T. K. (2006). *Evolutionary cognitive neuroscience*. Cambridge, MA: MIT Press.

Poldrack, R. A. (2008). The role of fMRI in cognitive neuroscience: Where do we stand? *Curr. Opin. Neurobiol.*, *18*(2), 223–227.

Price, G. R., Holloway, I., Vesterinen, N., Rasanen, P., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Curr. Biol.*, *17*, R1023–R1024.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. USA*, *98*(2), 676–682.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Dev.*, *79*, 375–394.

Ramus, F. (2006). Genes, brain, and cognition: A roadmap for the cognitive scientist. *Cognition*, *101*(2), 247–269.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, *27*, 169–192.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nat. Rev. Neurosci.*, *8*(9), 657–661.

Small, G. W., Bookheimer, S. Y., Thompson, P. M., Cole, G. M., Huang. S.-C., Kepe, V., & Barrio, J. R. (2008). Current and future uses of neuroimaging for cognitively impaired patients. *Lancet Neurol.*, *7*(2), 161–172.

Stern, W. M., Tormos, J. M., Press, D. Z., Pearlman, C., & Pascual-Leone, A. (2007). Antidepressant effects of high and low frequency repetitive transcranial magnetic stimulation to the dorsolateral prefrontal cortex: A double-blind, randomized, placebo-controlled trial. *J. Neuropsychiatry Clin. Neurosci.*, *19*(2), 179–186.

Stevens, C., & Neville, H. (2008). *Experience shapes human brain development and function: A framework for planning interventions for children at-risk for school failure.* Paper presented at AAAS meeting, Boston, MA.

Tovino, S. A. (2007). Functional neuroimaging and the law: Trends and directions for future scholarship. *Am. J. Bioeth.*, *7*, 44–56.

Varma, S., McCandliss, B. D., & Schwartz, D. L. (2008). Scientific and pragmatic challenges for bridging education and neuroscience. *Educ. Res.*, *37*, 140–152.

CONTRIBUTORS

Addis, Donna Rose Department of Psychology, University of Auckland, Auckland, New Zealand

Adolphs, Ralph California Institute of Technology, Pasadena, California

Aminoff, Elissa M. Department of Psychology, University of California, Santa Barbara, California

Andersen, Richard A. Computation and Neural Systems, Division of Biology, California Institute of Technology, Pasadena, California

Angelaki, Dora E. Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri

Arellano, Jon I. Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut

Arzi, Anat Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel

Aziz-Zadeh, L. Brain and Creativity Institute and The Department of Occupational Therapy, University of Southern California, Los Angeles, California

Badre, David Departments of Psychology and Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island

Baggio, Giosuè Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

Balduzzi, David Department of Psychiatry, University of Wisconsin, Madison, Wisconsin

Balslev, Daniela School of Psychology, University of California, Santa Barbara, California

Barlow, Horace Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom

Bavelier, Daphine Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York

Beck, Diane M. Department of Psychology, University of Illinois, Urbana-Champaign, Champaign, Illinois

Beer, Jennifer S. University of Texas at Austin, Austin, Texas

Bizzi, Emilio McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

Block, Ned Departments of Philosophy and Psychology; Center for Neural Science, New York University, New York, New York

Blumstein, Sheila E. Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island

Borroni, Paola University of Milano Medical School, Milano, Italy

Brainard, David H. Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania

Breunig, Joshua Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut

Bringcat, Scott Massachusetts Institute of Technology, Boston, Massachusetts

Brosch, Tobias Department of Psychology, University of Geneva; Swiss National Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Bruer, John T. James S. McDonnell Foundation, St. Louis, Missouri

Bryan, Ronald E. California Institute of Technology, Pasadena, California

Buckner, Randy L. Department of Psychology, Center for Brain Sciences, Harvard University, Cambridge, Massachusetts; Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts; Howard Hughes Medical Institute, Cambridge, Massachusetts

Bunge, Sylvia A. Helen Wills Neuroscience Institute and Department of Psychology, University of California at Berkeley, Berkeley, California

Burr, David Department of Psychology, University of Florence, Italy; School of Psychology, University of Western Australia, Perth, Australia

Cain, Christopher New York University, Center for Neural Science, New York, New York

Caplan, David Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts

Caramazza, Alfonso Department of Psychology, Harvard University, Cambridge, Massachusetts; Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy

Carroll, Joseph Department of Ophthalmology, Medical College of Wisconsin, Milwaukee, Wisconsin

Chalupa, Leo M. Department of Ophthalmology and Vision Science, School of Medicine and Department of Neurobiology, Physiology and Behavior, College of Biological Sciences, University of California, Davis, California

Chen, Yuzhi Department of Psychology and Center for Perceptual Systems, University of Texas, Austin, Texas

Chua, Elizabeth F. Center for Neuroscience, University of California, Davis, California

Cloutier, Jasmin Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge; Department of Psychology, Tufts University, Medford, Massachusetts

Cohen, Laurent AP-HP, Hôpital de la Salpêtrière, Department of Neurology, Paris; Université Paris VI, Faculté de Médecine Pitié-Salpêtrière, Paris; INSERM UMRS 975, Centre de Recherche de l'ICM, Paris, France

Connor, Charles E. Johns Hopkins University, Baltimore, Maryland

Corbetta, Maurizio Departments of Neurology, Radiology, and Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri

Crookes, Kate Department of Psychology, Australian National University, Canberra, Australia

Cross, Emily S. Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Davis, F. Caroline Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire
DeAngelis, Gregory C. Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York
Dehaene, Stanislas INSERM, Cognitive Neuro-Imaging Unit, Gif sur Yvette; Collège de France, Paris, France
Delgado, Mauricio R. Department of Psychology, Rutgers University, Newark, New Jersey
De Weerd, Peter Faculty of Psychology and Neuroscience, Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands
Dilkina, Katia Department of Psychology, Stanford University, Stanford, California
Doron, Karl W. Department of Psychology, University of California, Santa Barbara, California
Drew, Trafton University of Oregon, Eugene, Oregon
Dum, Richard P. Center for the Neural Basis of Cognition, Systems Neuroscience Institute and the Department of Neurobiology, University of Pittsburgh, Pittsburgh, Pennsylvania
Dye, Matthew W. G. Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York
Fisher, Simon E. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
Fitch, W. Tecumseh School of Life Sciences, University of Vienna, Vienna, Austria
Fogassi, Leonardo Dipartimento di Psicologia and Dipartimento di Neuroscienze, Università di Parma, Parma, Italy
Fries, Pascal Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands
Friston, Karl J. Wellcome Centre for Imaging Neuroscience, University College, London, United Kingdom
Funk, Chadd M. Department of Psychology, University of California, Santa Barbara, California
Gallese, Vittorio Dipartimento de Neuroscienze, Università di Parma, Parma, Italy
Ganis, Giorgio Massachusetts General Hospital, Charlestown, Massachusetts
Gazzaniga, Michael S. SAGE Center for the Study of Mind and Department of Psychology, University of California, Santa Barbara, California
Geisler, Wilson S. Department of Psychology and Center for Perceptual Systems, University of Texas, Austin, Texas
Gelstien, Shani Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
Gilbert, Charles D. The Rockefeller University, New York, New York
Gil-da-Costa, Ricardo The Salk Institute for Biological Studies, San Diego, California
Glimcher, Paul W. Professor of Neural Science, Economics, and Psychology; Director, Center for Neuroeconomics; New York University, New York, New York
Goebel, Rainer Faculty of Psychology and Neuroscience, Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands; Netherlands Institute for Neuroscience (NIN), an Institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), Amsterdam, The Netherlands
Grafton, Scott T. UCSB Brain Imaging Center, Department of Psychology, University of California Santa Barbara, Santa Barbara, California
Graybiel, Ann M. Department of Brain and Cognitive Sciences and the McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts
Green, C. Shawn Department of Psychology, University of Minnesota, Minneapolis, Minnesota
Greene, Joshua D. Department of Psychology, Harvard University, Cambridge, Massachusetts
Griffiths, Timothy D. Institute of Neuroscience, Newcastle University, Newcastle upon Tyne; Wellcome Centre for Imaging Neuroscience, University College, London, United Kingdom
Gu, Yong Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri
Guerin, Scott A. Department of Psychology, University of California, Santa Barbara, California
Haddad, Rafi Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
Hagoort, Peter Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen; Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
Hall, Julie L. Department of Psychology, University of Michigan, Ann Arbor, Michigan
Hariri, Ahmad R. Department of Psychology and Neuroscience, Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina
Heatherton, Todd F. Department of Psychological and Brain Sciences, Center for Cognitive Neuroscience, Datmouth College, Hanover, New Hampshire
Heinze, Hans-Jochen Department of Neurology, Otto von Guericke University; Leibniz Institute for Neurobiology, Magdeburg, Germany
Hickock, Gregory Center for Cognitive Neuroscience, University of California, Irvine, California
Hillyard, Steven A. Department of Neuroscience, University of California, San Diego, California
Holyoak, Keith J. Department of Psychology, University of California, Los Angeles, California
Hopf, Jens-Max Department of Neurology, Otto von Guericke University; Leibniz Institute for Neurobiology, Magdeburg, Germany
Horng, Sam Department of Brain and Cognitive Sciences, The Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts
Huberman, Andrew D. Department of Neurobiology, Stanford University School of Medicine, Stanford, California
Ivry, R. B. Department of Psychology, University of California Berkeley, Berkeley, California
Jordan, Kerry E. Department of Psychology, Utah State University, Logan, Utah
Judaš, Miloš Croatian Institute for Brain Research, School of Medicine, University of Zagreb, Zagreb, Croatia
Kanwisher, Nancy McGovern Institute for Brain Research and Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, Massachusetts
Karnath, Hans-Otto Center for Neurology, University of Tübingen, Tübingen, Germany

KASTNER, SABINE Department of Psychology, Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey
KENSINGER, ELIZABETH A. Department of Psychology, Boston College, Chestnut Hill; Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, Massachusetts
KHAN, REHAN Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
KIDD, JR., GERALD Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts
KNOWLTON, BARBARA J. Department of Psychology, University of California, Los Angeles, California
KOCH, CHRISTOF Division of Biology, California Institute of Technology, Pasadena, California
KOECHLIN, ETIENNE Institut National de la Santé et de la Recherche Médicale; Université Pierre et Marie Curie; Ecole Normale Supérieure, Paris, France
KOENIGS, MICHAEL University of Wisconsin, Madison, Wisconsin
KOSSLYN, STEPHEN M. Harvard University, Cambridge, Massachusetts
KOSTOVIĆ, IVICA Croatian Institute for Brain Research, School of Medicine, University of Zagreb, Zagreb, Croatia
KRAKAUER, JOHN W. The Motor Performance Laboratory, Department of Neurology, Columbia University College of Physicians and Surgeons, New York, New York
KUHL, BRICE A. Department of Psychology, Stanford University, Stanford, California
KUHL, PATRICIA K. Institute for Learning and Brain Sciences, University of Washington, Seattle, Washington
KUMAR, SUKHBINDER Institute of Neuroscience, Newcastle University, Newcastle upon Tyne; Wellcome Centre for Imaging Neuroscience, University College, London, United Kingdom
LANDAU, AYELET N. Department of Psychology, University of California, Berkeley, California
LAPID, HADAS Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
LAU, HAKWAN Columbia University, New York, New York
LEDOUX, JOSEPH E. New York University, Center for Neural Science, New York, New York
LEIBERG, SUSANNE Laboratory for Social and Neural Systems Research, Institute for Empirical Research in Economics, University of Zurich, Zurich, Switzerland
LI, WU Beijing Normal University, Beijing, China
LUCK, STEVEN J. Department of Psychology and Center for Mind and Brain, University of California, Davis, California
MACKEY, ALLYSON P. Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, California
MACKNIK, STEPHEN L. Barrow Neurological Institute, Phoenix, Arizona
MANGUN, GEORGE R. Center for Mind and Brain, University of California, Davis, California
MARTIN, ALEX Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, Maryland
MARTINEZ-CONDE, SUSANA Barrow Neurological Institute, Phoenix, Arizona
MAUNSELL, JOHN H. R. Department of Neurobiology, Harvard Medical School, Howard Hughes Medical Institute, Boston, Massachusetts
MCCLELLAND, JAMES L. Department of Psychology, Stanford University, Stanford, California
MCKONE, ELINOR Department of Psychology, Australian National University, Canberra, Australia
MCMAINS, STEPHANIE A. Department of Psychology, Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey
MILLER, MICHAEL B. Department of Psychology, University of California, Santa Barbara, California
MINK, JONATHAN W. Departments of Neurology, Neurobiology and Anatomy, Brain and Cognitive Sciences, and Pediatrics, University of Rochester School of Medicine and Dentistry, Rochester, New York
MITCHELL, JASON P. Department of Psychology, Harvard University, Cambridge, Massachusetts
MOLNAR-SZAKACS, ISTVAN Tennenbaum Center for the Biology of Creativity, Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California
MONTASER-KOUHSARI, LEILA Department of Psychology, New York University, New York, New York
MORICEAU, STEPHANIE Emotional Brain Institute, Nathan Kline Institute and Child and Adolescent Psychiatry, New York University Langone Medical Center, Orangeburg, New York
MORRONE, CONCETTA Department of Physiological Sciences, University of Pisa, and Scientific Institute Stella Maris, Pisa, Italy
MOSER, EDVARD I. Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, Norwegian University of Science and Technology, Trondheim, Norway
MOVSHON, J. ANTHONY Center for Neural Science, New York University, New York, New York
MULLIKEN, GRANT H. Computational and Neural Systems, California Institute of Technology, Pasadena, California
MUSSA-IVALDI, FERDINANDO A. Department of Physiology, Northwestern University Medical School, Chicago, Illinois
NADER, KARIM Psychology Department, McGill University, Montreal, Quebec, Canada
NEVILLE, HELEN Department of Psychology, University of Oregon, Eugene, Oregon
OLOFSSON, JONAS K. Department of Psychology, Umeå University, Umeå, Sweden; Psychology Department, Stockholm University, Stockholm, Sweden
OVERATH, TOBIAS Wellcome Centre for Imaging Neuroscience, University College, London, United Kingdom
PASCUAL-LEONE, ALVARO Berenson-Allen Center for Noninvasive Brain Stimulation, Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts
PASUPATHY, ANITHA University of Washington, Seattle, Washington
PATTERSON, KARALYN MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom
PHELPS, ELIZABETH A. Department of Psychology and Neural Science, New York University, New York, New York
PREUSS, TODD M. Division of Neuroscience and Center for Behavioral Neuroscience, Yerkes Primate Research Center, Emory University, Atlanta, Georgia

Quadflieg, Susanne School of Psychology, University of Aberdeen, Aberdeen, Scotland
Race, Elizabeth A. Neurosciences Program, Stanford University, Stanford, California
Raichle, Marcus E. Washington University School of Medicine, St. Louis, Missouri
Raineki, Charlis Emotional Brain Institute, Nathan Kline Institute and Child and Adolescent Psychiatry, New York University Langone Medical Center, Orangeburg, New York
Rakic, Pasko Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut
Ralph, Matthew Lambon Department of Psychology, University of Manchester, Manchester, United Kingdom
Ramus, Franck Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS, CNRS, DEC-ENS, Paris, France
Rangel, Antonio Humanities and Social Sciences, Computational and Neural Systems, California Institute of Technology, Pasadena, California
Rees, Geraint Institute of Cognitive Neuroscience and Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom
Richards, Virginia M. Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania
Ringach, Dario L. Department of Neurobiology and Psychology, Jules Stein Eye Institute, David Geffen School of Medicine, University of California, Los Angeles, California
Rizzolatti, Giacomo Dipartimento di Neuroscienze, Università di Parma, Parma, Italy
Robertson, Lynn C. Veterans Administration Research, Department of Psychology, and Helen Wills Neuroscience Institute, University of California, Berkeley, California
Rogers, Timothy T. Department of Psychology, University of Wisconsin, Madison, Wisconsin
Roth, Tania L. Department of Neurobiology and the Evelyn F. McKnight Brain Institute, University of Alabama at Birmingham, Birmingham, Alabama
Saron, Clifford D. Center for Mind and Brain, University of Califoria, Davis, California
Schacter, Daniel L. Department of Psychology, Harvard University, Cambridge, Massachusetts; Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts
Schadmehr, Reza Laboratory for Computational Motor Control, Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, Maryland
Schiff, Nicholas D. Laboratory for Cognitive Neuromodulation, Department of Neurology and Neuroscience, Weill Medical College, Cornell University, New York, New York
Schiller, Daniela New York University, Center for Neural Science, New York, New York
Schoenfeld, Mircea A. Department of Neurology, Otto von Guericke University; Leibniz Institute for Neurobiology, Magdeburg, Germany
Seidemann, Eyal Department of Psychology and Center for Perceptual Systems, University of Texas, Austin, Texas
Sela, Lee Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
Shadmehr, Reza Laboratory for Computational Motor Control, Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, Maryland
Shapiro, Kevin A. Department of Psychology, Harvard University, Cambridge, Massachusetts; Department of Medicine, Children's Hospital, Boston, Massachusetts
Shrager, Yael Department of Neurosciences, University of California, San Diego, La Jolla, California. Now at Department of Psychology, Harvard University, and Howard Hughes Medical Institute, Cambridge, Massachusetts
Shulman, Gordon L. Departments of Neurology, Radiology, and Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri
Simoncelli, Eero P. Center for Neural Science and Courant Institute of Mathematical Sciences, New York University, New York, New York
Singer, Tania Laboratory for Social and Neural Systems Research, Institute for Empirical Research in Economics, University of Zurich, Zurich, Switzerland
Sobel, Noam Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
Somerville, Leah H. Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University, New York, New York
Squire, Larry R. Veterans Affairs Healthcare System, San Diego, California; Department of Psychiatry, Department of Neurosciences, Department of Psychology, University of California, San Diego, La Jolla, California
Stephan, Klaas E. Wellcome Centre for Imaging Neuroscience, University College, London, United Kingdom
Stevens, Courtney Willamette University, Salem, Oregon
Strick, Peter L. Veterans Affairs Medical Center; Center for the Neural Basis of Cognition, Systems Neuroscience Institute and the Department of Neurobiology, University of Pittsburgh, Pittsburgh, Pennsylvania
Sullivan, Regina M. Emotional Brain Institute, Nathan Kline Institute and Child and Adolescent Psychiatry, New York University Langone Medical Center, Orangeburg, New York
Summerfield, Christopher Oxford University, Oxford, United Kingdom
Sur, Mriganka Department of Brain and Cognitive Sciences, The Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts
Suzuki, Wendy A. Center for Neural Science, New York University, New York, New York
Sy, Jocelyn L. Department of Psychology, University of California, Santa Barbara, California
Sylvester, Chad M. Departments of Neurology, Radiology, and Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri
Thompson, William L. Harvard University, Cambridge, Massachusetts
Todorov, Emanuel Department of Cognitive Science, University of California San Diego, San Diego, California
Tononi, Giulio Department of Psychiatry, University of Wisconsin, Madison, Wisconsin
Treisman, Anne Psychology Department, Princeton University, Princeton, New Jersey
Uddin, Lucina Q. Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, California
Von Kriegstein, Katharina Wellcome Centre for Imaging Neuroscience, University College, London, United Kingdom

VUILLEUMIER, PATRIK Department of Neuroscience, University Medical Center, Geneva; Swiss National Center for Affective Sciences, University of Geneva, Geneva, Switzerland

WAGNER, ANTHONY D. Department of Psychology and Neurosciences Program, Stanford University, Stanford, California

WALSH, BONG J. Center for Mind and Brain, University of California, Davis, California

WANDELL, BRIAN A. Department of Psychology, Stanford University, Stanford, California

WEISBROD, AHARON Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel

WHALEN, PAUL J. Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire

WHITAKER, KIRSTIE J. Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, California

WHITLOCK, JONATHAN R. Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, Norwegian University of Science and Technology, Trondheim, Norway

WILLEMS, ROEL M. Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

WILLIAMS, DAVID R. Center for Visual Science, University of Rochester, Rochester, New York

WOMELSDORF, THILO Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

WRIGHT, BEVERLY A. Department of Communication Sciences and Disorders and Interdepartmental Neuroscience Program, Northwestern University, Evanston, Illinois

YAMADA, MIKIKO Department of Neuropsychiatry, School of Medicine, Kyoto University, Kyoto, Japan

YAMANE, YUKAKO Riken Brain Science Institute, Saitama, Japan

YESHURUN, YAARA Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel

YOON, GEUNYOUNG Center for Visual Science, University of Rochester, Rochester, New York

ZHANG, YUXUAN Department of Communication Sciences and Disorders and Interdepartmental Neuroscience Program, Northwestern University, Evanston, Illinois

INDEX

A

Abulia, 568, 573
Acetylcholine, 1140
Acetylcholinesterase (AChE), laminar shifts in, 34
Across-stimulus generalization, in auditory training, 357–358
Across-task generalization, in auditory training, 355–357
Actin polymerization, 146–147
Action
  in intention understanding, 635–636
  in learning and learning transfer, 160–161
  object concepts in support of, 1032–1035
Action-based choice, 1081
Action planning
  central nervous system and, 541–543, 547
  hierarchical nature of, 642–644
  motor primitives in, 541–543
Action semantics, 648–650
Actions
  fear, neural basis of, 914–918
  nature of, 906
  in Pavlovian conditioning, 906–909, 914–918
Actual code, 420, 432
Actual decoder, 420, 432
Actual encoder, 420
Ad auctions, 1250–1251
Adaptation
  aftereffects in face recognition, 468
  defined, 436–437
  eye movements and, 517–518
  fMRI, 715
Adaptive control, basal ganglia function and, 375–377
Adolescence
  cognitive development in, 73–82
    cognitive control, 73, 76–77
    current and future directions for research, 81–82
    fluid reasoning, 77–81
    structural brain development, 73–74
    working memory development, 73, 74–76
  stress and, 893
Adoption studies, language impairment, 856
Adulthood, stress and, 893
Affective theory, 1093–1094
Afference, posterior parietal cortex (PPC) and, 600–601
Aggression, violent video games and, 155
Aging process
  aerobic exercise and cognitive processes, 157
  Alzheimer's disease, 147, 270, 1009–1010, 1237
  episodic memory deficits, 753
  relational reasoning and, 1008–1011
Agnosia, 792
Agrammatic aphasia, 808
Agraphia, 799
Akaike information criterion (AIC), 378, 379
Akakievitch, Akakhi, 310
Alexia, 260, 794, 799
Alexithymia, 979
Alhazen, 511, 514
Allelic variation, 89–90
Altered states of consciousness, 1139
Alzheimer's disease, 147, 270, 1009–1010, 1237
American Association of Anatomists, 9
American Sign Language (ASL), 169–170
Amnesia
  in blocking reconsolidation, 694–695
  cue-induced, 693–694
  experimental, 696–697
  future-event simulation and, 752–754
  HM (patient), 656, 659, 675
  learning rewarding nature of sensory states, 592–593
  medial temporal lobe function
    habit learning studies, 679–680
    retrograde amnesia, 684–685
    working memory studies, 678–679
  recovery from, 696–697
  source, 1118
  time-dependent behavioral impairment, 694–695
AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole proprionate) receptors
  in fear reaction, 912
  synaptic plasticity and, 111–113, 118–119
AMPAR endocytosis, 101
Amygdala
  context effects and, 935–942
    context conditioning, 939–940
    experimental context, 941–942
    facial expression, 935–939
    fMRI activation to specific stimuli, 936–937, 941–942
  and domain-specific neural circuitry, 1036–1038
  in emotion regulation, 966
    nonconscious processing of emotion, 1186
    social cognition, 1185
  in fear response, 1095–1096
    conditioned fear actions, 914–918
    conditioned fear reactions, 891–896, 911–913, 930
    infant, 891–896
    parallel processing, 913–914
  memory for visual details of emotional stimuli, 726–727
  modulation of hippocampal consolidation, 729–730
  moral judgment and, 988, 989–990
  organization of, 909–911
  overview, 1095–1097
  in processing emotional information, 928–929, 931, 932
    fMRI activation to specific stimuli, 936–937, 941, 942
    imaging genetics studies, 947–948
  in retrieval of emotional memories, 731–733
  in self-regulation, 957
  in threat detection, 958
  valuation in decision making, 1097
Analogical reasoning, 78, 79, 1008, 1010, 1012–1013
Anger, facial expression and amygdala response, 935
Animal studies. *See* Birds; Cats; Comparative approach; Ferrets; Fish; Frogs; Monkeys; Neural populations in primate cortex; Rodents
Anisomycin, 698
Anisotropy, 1127–1128
Anomia, 260
Anoxia, 1128–1130
Anterior area (AI), empathy and, 975–982
Anterior cingulate cortex (ACC)
  in conditioned fear reactions, 917, 918
  conflict-control model, 252–257
  in decision making, 1023–1027
  in emotion regulation, 963, 964, 965–970
  empathy and, 975–981
  moral judgment and, 988–991, 992, 993–994
  regulation of control, 709
  reward/value signals in limbic system, 228
  in threat detection, 957–958
Anterior inferior parietal area (AIP), mirror neuron system and, 626–633, 634–635
Anterior intraparietal sulcus (aIPS), goal representation and, 644–645, 647
Anterior olfactory nucleus, 326
Anterior rostral MFC, 1026–1027
Anterior temporal lobes (ATL), in semantic cognition, 1061–1064
Anterograde alteration, of synaptic plasticity, 114–116
Anticipation, 962
  anticipatory signals, 219–223
  movement intention and, 599–601
Antidepictive theories, 1243
Antisocial personality disorder (APD), moral judgment and, 988
Aphasia, 148, 259–260, 1033, 1236
  double dissociation in processing, 767–768, 780, 781
  following stroke, 259–260, 810
  morphological deficits in, 779, 780
  phonological processing in, 767–768
  semantic, 1063–1064
  syntactic deficits and, 808–813

Apoptosis (programmed cell death), cross-species comparisons of cortical development, 16
Apperceptive agnosia, 792
Appetitive conditioning research, 914
Apraxia, 599, 648–649
  conceptual, 649
  ideomotor, 649
Aquatic cetaceans, in evolutionary biology, 52
Arcuate fasciculus (AF), connectivity studies, 57–58, 261–262, 263
Areal domain, neocortical development, 7, 18–19
Aristotelian philosophy, 49
Aristotle, 1116
Arousal
  in learning and learning transfer, 161
  states of consciousness and, 1139–1140
Artificial grammar learning, 880
Asperger syndrome, empathy and, 979
ASPM (abnormal spindlelike microcephaly associated), 59
Assessment. *See names of specific assessment instruments*
Assimilation, in color vision, 387–388
Association studies
  basal ganglia function and, 576
  nature of, 856
Associativity
  in conceptual representations, 1032
  in memory formation, 113–114, 119, 121–122
  in valuation process, 1087
Asynchrony detection, in relative-learning tasks, 359
Ataxia, 603
Athletic domain
  aging process and, 157
  complex learning environments, 156–158
  experience-dependent learning, 155, 157
Atonia, 1139
Attachment, infant
  fear learning and, 891–896
  maternal separation/deprivation in, 895
  nature of, 889–890
  neonatal handling, 895–896
  neural circuitry and, 890–891
  odor learning in, 889–890
  rearing environment alteration, 895
Attention. *See also* Cognitive (executive) control; Selective attention; Visual attention
  attention training for children, 173–175, 176
  attentional control processes, 185–186, 251–252
  basal ganglia function and, 573–575
  behavioral phenomena of, 189–200
    attentional bottleneck, 282–283
    consciousness as, 199–200
    control of attention, 198–199
    feature binding, 197–198, 274–276
    focused versus distributed attention, 199
    limits on attention, 189–191
    in perceptual processing, 926–928
    selection process, 191–193
    timing, 193–197
  consciousness versus, 199–200, 1141–1143
  control of, 186, 198–199
    integrating with conflict detection, 251–257
    Sylvian fissure, 186, 263–264
  defining, 185, 219
  effect of emotion on allocation of, 726–727
  emotion and, 925–932
    automaticity, 929–931
    behavioral effects on perception, 926–928
    cultural factors, 931–932
    current behavior goals, 932
    neural circuits underlying attention, 928–929
    neural processing in perception, 925–926
    personality factors, 931–932
  event-related potentials (ERPs) in, 170, 176, 186, 193, 196–197, 199, 254–255
  feedback in visual, 1165–1175
  focused versus distributed, 199
  Hillyard principle, 170
  limits of, 189–191
    behavioral coherence, 191
    disappearance of, 191
    general resources, 190
    structural interference, 189–190
  method of selection, 192–193
    changes of tuning or selectivity, 193
    facilitation, 192–193
    inhibition, 193
  microcircuitry of, 186
  orientation of spatial, 795–796
  plasticity in development of, 170–173
  priming and, 196, 197, 199
  stages of, 193–197
    attentional blink, 196, 197, 199, 1153, 1209
    early versus late, 193–197
    explicit processing and, 196, 197
    implicit processing and, 195–197
    psychological refractory period, 194, 195
    reverse hierarchy theory (Hochstein and Ahissar), 194
  top-down control signals, 186–187, 211–212, 213–214, 220
  video games and, 157
Attention-deficit/hyperactivity disorder (ADHD), 82, 147, 176, 865
  basal ganglia function and, 568, 574
Attention Network Test (ANT), 192
Attentional blink, 196, 197, 199, 1153, 1209
Attentional bottleneck, 282–283
Attentional control signals. *See also* Cognitive (executive) control
  conflict and, 252–253
  nature of, 251
Attentional dyslexia, 797
Attentional load theory, 206–207
Attentional processing, in selective attention, 293–298
  feature-selective modulation of rhythmic synchronization, 297–298
  gamma-band synchronization, 293–295
  inter-areal synchronization, 298–299
  preparatory attentional states, 295
  temporal expectancies of target processing, 296–297
Attractive bias, 531
Attractor networks, 121–123
Auditory cortex
  neuroplasticity in, 147, 167–169
  in perceptual learning, 132–133, 353–363
  peripheral auditory attention, 170–171
  rewiring vision into auditory pathway, 95–98, 131, 166–167
  visual deprivation and auditory development, 167–169, 321
Auditory filter, 345
Auditory masking, 343–351
  energetic, 344–345, 349, 351
  informational, 345–351
    described, 345–348
    simultaneous multitone maskers and, 345–348
    speech recognition process, 348–351
  perceptual, 348–349
  simultaneous multitone maskers, 345–348
Auditory-motor integration networks, verbal working memory in, 771
Auditory pathway, rewiring vision into, 95–98, 131, 166–167
Auditory processing, 132–133, 353–363
  characteristics of, 358–362
    across-task generalization, 355–357
    learning on trained condition, 354–355
  in fear conditioning, 913
  neural processes in
    described, 353–358
    neural underpinnings of perceptual learning, 362–363
  object analysis, 367–380
    concept of auditory object, 367–368
    fMRI studies in, 371–379
    manipulation of natural stimuli, 368–370
    natural stimuli, 368–370
    stimuli based on sequence of objects, 370–371
    synthetic stimuli for, 370
Autism, 147, 166
  empathy and, 979
  motor cognition and, 636–638
  phenomenal consciousness in, 1118
Autism spectrum disorder (ASD), early language acquisition, 845–847, 848
Autobiographical memory, 684–686, 726, 731, 753
Autobiographical Memory Interview (AMI), 686, 687, 753
Automaticity, of emotional processing, 929–932
Autosomal dominant retinitis pigmentosa (adRP), 388–389
Avians. *See* Birds
Avoidance conditioning, 116, 907–909

Awareness. *See also* Consciousness
declarative memory and, 686–687
detecting, following brain injury, 1125–1127
self-awareness, 954–955
visual. *See* Visual awareness
Axel, Richard, 324–325

B

Back-propagation of error, 1050
Balanced Emotional Empathy Scale (BEES), 978–979
Balint's syndrome
characteristics of, 269, 270–273, 798
following stroke, 198
loss of feature-based attention, 271–272
loss of object-based attention, 270–271
Bartlett, Frederic, 751
Basal ganglia
in cognition, 565–579
anatomical perspectives, 565–568
chunking of action repertoires, 577–579
clinical perspectives, 568
hypotheses of function, 568–577
functions, 553–554, 568–577, 709
pathology of, 560–561
virus tracing studies, 553–561
Baseball, experience-based learning in, 155, 157
Basic fibroblast growth factor (bFGF), 30
Basketball, complex learning environments, 156–157
Basolateral complex (BLA), in fear learning/conditioning, 911, 913, 915–917, 919
Basolateral nuclei, 911
Basomedial nuclei, 911
Bayesian information criterion (BIC), 378, 379, 380
Bayesian statistics, 396–406
basic ideas, 396–397
color constancy and, 398–401
cone mosaic and, 401–404
empirical estimation, 529
ideal Bayesian observer, 420–421, 530
inference and, 159
duality with optimal control, 614–622
linear estimator, 528
maximum a posteriori (MAP) estimator, 528–529, 532
models in, 397–398, 404–406
optimal estimation and, 527–529
perceptual Bayesianism, 530–531
quadratic error (least squares) solution, 527–528
unsupervised regression, 529
Behavioral measures
of attention, 189–200
attentional bottleneck, 282–283
consciousness as, 199–200
control of attention, 198–199
feature binding, 197–198, 274–276
focused versus distributed attention, 199
limits on attention, 189–191
selection process, 191–193
timing, 193–197
of facial recognition, 467–470, 473–474, 477–478
in goal-directed choice
multiple behavioral controllers, 1078–1079
simple binary stimulus, 1076–1077, 1079–1081
of multisensory integration, 505–506
Belief networks, 614–621
Belief propagation, 621
Bellman equation, 617
Bereitschaftspotential (BP), spontaneous motor initiation, 1191
Beta-band synchronization, 298–299
Biased competition theory, 198–199, 705–706
Bibliometric studies, 1221–1222
Bilingualism, infant, 848–849
Bimanual movements, goal representation in, 645–647
Binocular correlation, in depth perception, 489–490
Binocular disparity, in depth perception, 483, 484–489
Binocular rivalry, 1143, 1156, 1170–1176, 1209
Biolinguistics, 875–876
Biological theory of consciousness
described, 1112–1113
explanatory gap and, 1114
self in, 1118–1119
Birds
cross-species comparisons of neocortical development, 8
mechanisms underlying vocal learning, 878–879
Blind spots, filling-in, 436–437
Blindness
attention and, 170–171
consciousness as requirement for exclusion, 1194–1195
inattentional, 195–196
metacognitive masking and, 1198, 1199
nonconscious processing of emotion, 1186
sensory substitution and, 167–169, 321
visual area involvement in nonvisual tasks, 168–169
Bliss, Tim, 109–111
Blue cone monochromacy, 390
BOLD fMRI. *See under* Functional magnetic resonance imaging (fMRI)
Bottom-up influences, in selective attention, 212–214
attentional processing, 298–299
relation to top-down influences, 213–214
scene segmentation, 212–213
visual salience, 212
Boulder Committee model, 10, 11, 12
Bound functional morphemes (inflections), 777
Boundary processing
in controlling spread, 439–441
in object recognition process, 455–460, 463–464
in surface perception, 436–437, 439–441, 442–443
Brachium of the IC (BIC), 95
Bradykinesia, 571, 573
Bradyphrenia, 573
Brain damage. *See also* Lesion studies; Stroke
moral judgment and, 987–988
parahippocampal region, 659–660
recovery of consciousness after brain injury, 1123–1134
central thalamic deep-brain stimulation, 1131–1134
circuit mechanisms underlying forebrain function, 1128–1130
detecting awareness in absence of behavioral responsiveness, 1125–1127
disorders of consciousness overview, 1124–1125
functional recovery and white matter structural changes, 1127–1128
future directions, 1133–1134
striatal, 593–594
virtual, 1005
Brain-derived neurotrophic factor (BDNF), 100, 145, 146–147
Brain specialization, 53–61
comparative genomics, 58–61
comparative histology, 54–55
comparative neuroimaging, 55–58
Brain surgery, plasticity as opportunity for intervention, 148
Brain volume, 746
BrdU, as marker for DNA synthesis, 8
Broadbent, W. H., 1032–1033
Broadly congruent, 629
Broca, Paul, 1235
Broca's area
aphasia and, 741, 773, 808, 809, 811, 1236, 1237
comparative neuroimaging, 57
connectivity studies, 261
in early language acquisition, 847–848
in human brain specialization, 53
morphological processing and, 782
neuronal circuitry of frontal lobe, 39
in semantic unification network, 831
Brodmann's area, 1021, 1236
Buck, Linda, 324–325
Burst neurons, 1089–1090
Bush, George H. W., 1260

C

Cajal-Retzius cells, 33, 37, 42
Calbindin, in human brain specialization, 55, 56
Calcium
cross-species comparisons of cortical development, 14
in long-term potentiation (LTP) process, 146
in postsynaptic terminal, 111, 112
Calcium/calmodulin-dependent kinase II (CaMKII), 116
Calretinin, in comparative histology, 54
Cancer, 858
Capacity for culture, 875
Caring, moral judgment and, 988–989

Cascade model, 1021–1023
described, 1021–1022
Caspases, cross-species comparisons of cortical development, 16
Cat-301, 55, 56
Cats
sensory substitution in, 321
visual cortex
formation of eye-specific inputs, 68
temporal code in perceptual learning, 136
Caud, caudate nucleus, neuronal circuitry of frontal lobe, 31
Causal manipulation, in depth perception, 492–493
Cell divisions, cross-species comparisons of cortical development, 17
Cellular (synaptic) consolidation, 692
Central and anterior IT (CIT/AIT) neural tuning, 461–465
Central nervous system (CNS)
mechanical basis for compositionality, 547–548
planning and, 541–543, 547
Central pattern generators (CPGs), 545–546
Centrosome, cross-species comparisons of cortical development, 14
Cerebellum
functions, 553–554
in motor learning and control
construction of internal models, 592
predicting sensory consequences of motor commands, 591–592
pathology of, 560–561, 592
virus tracing studies, 553–561
Cerebral cortex
cross-species size comparisons, 15–17
development of, 314–316
first appearance, 315–316
hierarchical function, 314–316
symmetry and, 316–317
virus tracing studies, 559
Cetaceans
comparative histology, 55
in evolutionary biology, 52
Childhood. *See also* Developmental stages
attention training in, 173–175
attentional modulation in, 171
autism in, 147, 166, 636–638, 979, 1118
cognitive development in, 73–82
cognitive control, 73, 76–77
current and future directions for research, 81–82
fluid reasoning, 77–81
structural brain development, 73–74
working memory development, 73, 74–76
differential development and, 1048
early childhood stage
defined, 30
neuronal circuitry of frontal lobe, 30, 31, 41–44
early language acquisition, 837–851
infant lexicon, 843–844
neuroscience-based measures, 837–839
phonetic learning, 839–841
second-language learning, 842
sentence processing, 844–845
word learning, 842–843
face recognition in, 473–478
phenomenal consciousness during, 1117–1118
relational reasoning during, 1005
Chimpanzees. *See* Monkeys
Choice. *See* Consciousness, volition and function of; Decision making; Goal-directed choice
Choice probabilities (CP), in depth perception, 490–492
Chomsky, Noam, 806, 807–808, 837, 855
Chondroitin-sulfate proteoglycans (CSPGs), in ocular dominance plasticity, 101
Chorea, 572
Chronic neuropathic pain syndromes, 147
Chunking, basal ganglia function and, 577–579
Cinematographic vision, 1145
Class I major histocompatibility complex (Class I MHC), 147
Closed loop control, 615–617
CNTNAP2, 862, 863–864
Coarse depth discrimination, 489–493
Cognitive (executive) control
basal ganglia in, 567
conflict and, 252–257
in decision making, 1021
defined, 705
development of, 76–77, 82, 705–709, 714–719, 1019–1027
in emotion regulation, 969
functions within, 76
memory and, 705–720
interactions, 709–714
mnemonic prediction, 714–719
mnemonic suppression, 714–719
prefrontal cortex in, 76–77, 82, 705–709, 714–719
theoretical basis, 705–709
nature of, 73
reliance on, 1247
types, 1021
Cognitive evolution. *See* Evolutionary biology
Cognitive function
activations in, 1067–1068
assumptions of pure insertion, 1067–1068
basal ganglia in, 565–579
anatomical perspectives, 565–568
chunking of action repertoires, 577–579
clinical perspectives, 568
hypotheses of function, 568–577
cost of intrinsic activity, 1068–1070
cross-species comparisons of neocortical development, 7–8, 20, 22–23
current and future directions for research, 81–82
fluid reasoning development, 77–81
in human brain specialization, 53–61
imagery debate, 1241–1245
levels of
algorithmic, 807–810
neural, 810–813
representational, 805–807
neuronal circuitry of frontal lobe, 29–44
organization of intrinsic activity, 1070–1071
specificity of learning, 154, 156
stages of, 29–44
structural brain development, 73–74
theory of cognition, 1241–1243
competence versus processing, 1241–1243
relevance of the brain, 1243–1244
working memory development, 74–76, 82
Cognitive neuroscience
ethics and, 1260
imagery debate, 1241–1245
landscape of, 1255–1261
integrating, 1258–1259
methodology, 1255–1256
society and, 1259–1261
theory, 1256–1258
of language, 1235–1239
mapping, 1221–1234
bibliometric studies, 1221–1222
journal citation maps, 1222–1230
topic maps, 1230–1233
origins of, 1221, 1248, 1255
purpose of, 1247
as term, 1221
Cognitive perspective taking. *See* Empathy
Color blindness, 389–390
Color vision, 383–392
Bayesian approaches to, 396–406
basic ideas, 396–397
color constancy and, 399–401
cone mosaic and, 401–404
models in, 397–398, 404–406
color constancy, 398–401
deficient, 388–392
blue cone monochromacy, 390
red/green deficiency, 389–390
rod monochromacy, 390–392
tritanopia, 388–389
fundamentals of, 395
normal, 383–388
L cone mosaic, 386–388
M cone mosaic, 386–388
S cone mosaic, 384–386
object concepts and, 1034–1035
visual plasticity and, 166, 167
Coma, 1124
Communication. *See also* Language processing
attention and stages of, 193–197
Comparative approach
differentiation level in, 52–53
to human brain specialization
comparative genomics, 58–61
comparative histology, 54–55
comparative neuroimaging, 55–58
model-animal approach versus, 52–53
nature of, 856–857
to neocortical development, 7–23
cognitive processing, 7–8, 20, 22–23
cortical architecture, 7
cortical size determinants, 15–17
neuronal cell migration, 10–14
onset of neurogenesis, 8
origins of cortical neurons, 8–9
protomap hypothesis, 18–19

radial unit hypothesis, 14–15
span of neurogenesis, 8
stages of cortical development in humans, 20–22
synaptic connections, 7, 19–20
transient embryonic zones, 9–10
to parahippocampal region
boundaries and nomenclature, 660–662
connectivity in, 662–669
damage, 659–660
defined, 660
entorhinal cortex, 659–660, 661, 662, 667–671
monkeys versus rats, 659–671
parahippocampal cortex, 659, 661, 665–666
perirhinal cortex, 659, 660–665
postrhinal cortex, 663, 666–667, 670
Competence, processing versus, 1241–1242
Competition
control pathways in basal ganglia, 569–573
in memory tasks, interference, 76, 710–719
in selective attention, 186
biased competition theory, 198–199, 705–706
bottom-up influences on, 212–214
neural basis of competition, 210–211
observing sensory biases, 219–223
top-down influences on, 211–212, 213–214
Complex genetic diseases, 858
Complex learning environments, 156–158
Complexes, in integrated information theory (IIT), 1205, 1207
Composite effect, in face recognition, 468
Compositionality, motor system, 543, 547–548
Concentration, 185. *See also* Attention
nature of, 185
Concepts
conceptual dualism, 1115
defined, 1115
Conceptual priming, 715–719
Conceptual representation, 1031–1042
neural foundations, 1032–1042
acting, 1032–1035
anterior regions of temporal lobe, 1040–1041
domain-specific neural circuitry, 1035–1041
feeling, 1032–1035
perceiving, 1032–1035
visual processing, 1032
object maps and, 1031–1032
property information, 1035–1038
Conceptual self, 954
Conditioned motivation, 909, 915–918
Conditioned response (CR)
conditioned fear reactions, 891–896, 911–918
parallel processing, 913–914
serial processing, 911–913, 914
in context conditioning, 939–940
in decision making, 1095–1096
nature of, 906–909
Conditioned stimulus (CS)
in context conditioning, 939–940
in decision making, 1095–1096
nature of, 906–909
Configuration space, 542–543
Conflict
attentional control systems and, 252–253
brain networks for cognitive processing, 253–255
cognitive control and, 252–257
Conflict adaptation, 1024–1025
Connectionism, 158–159
Connectivity
anatomical basis for, 437–438
anticipatory signals
defined, 219–223
dorsal frontoparietal attention network, 219–223
for spatial attention, 222–223
auditory object analysis, 375–379
cascade model and, 1023
functional connectivity
defined, 219–223
dorsal frontoparietal attention network, 219–223
by fMRI, 221
importance of, 1252
neuroimaging studies of, 55–58
in parahippocampal region, 662–669
perisylvian white matter, 260–262
statistical connectivity theory, 411–415
cortical maps, 413–414
described, 411–413
testing, 414–415
surface perception and, 437–438
Consciousness. *See also* Awareness
attention versus, 199–200, 1142–1143. *See also* Attention; Visual attention
coma, 1124
defined, 1137, 1181
development of, 316
emotion and, 1181–1188
conscious experience of emotion, 1181–1185
nonconscious processing of emotion, 1185–1188
in evolutionary biology, 1138–1139
locked-in state (LIS), 1124–1125
minimally conscious state (MCS), 1123, 1124, 1125–1127, 1130–1134, 1139
nature of, 752, 1137
neurobiology of, 1137–1147
arousal and states of, 1139–1140
attention versus, 1142–1143
forward versus feedback projections, 1145–1146
free will and, 1138
interaction versus, 1141
neuronal basis of perceptual illusions, 1143–1145
neuronal circuitry of frontal lobe, 41
neuronal correlates, 1140–1141
neuronal correlation of consciousness (NCC), 1140–1141, 1143–1145, 1172–1176
in other species, 1138–1139
quantum mechanics and, 1141
self-consciousness versus, 1141–1142
phenomena included in, 1137–1138
recovery after brain injury, 1123–1134
central thalamic deep-brain stimulation, 1131–1134
circuit mechanisms underlying forebrain function, 1128–1130
detecting awareness in absence of behavioral responsiveness, 1125–1127
disorders of consciousness overview, 1124–1125
functional recovery and white matter structural changes, 1127–1128
future directions, 1133–1134
separation of cognition and, 1112
specificity of, 1209–1212
theories of, 1111–1120, 1201–1216
biological, 1112–1113, 1114, 1118–1119
consciousness-of, 1115–1116, 1182
determining quality of consciousness, 1209–1217
determining quantity of consciousness, 1201–1209, 1210–1211
dualism and, 1114–1115
explanatory gap and, 113–1115
global workspace, 1111–1112, 1113, 1114, 1119
higher order thought (HOT), 1111, 1113, 1114, 1115–1118, 1119
machine consciousness and, 1119–1120
nature of concept, 1115
self and, 1118–1119
vegetative state (VS), 1124
visual awareness, 1151–1161
brain activity and, 1152
causal factors and pathologies of vision, 1158–1159
characterizing unconscious homunculus, 1152–1154
empirical ad theoretical integration, 1159–1160
impact of, 1155–1158
measuring, 1151–1152
unconscious vision and multivariate pattern analysis, 1154–1155
volition and function of, 1191–1199
conscious veto, 1193–1194
exclusion, 1194–1195
inhibition, 1194–1195
spontaneous motor initiation, 1191–1193
top-down cognitive control, 1195–1197
true function of consciousness, 1197–1199
Consciousness-of, nature of, 1115–1116, 1182
Consolidation theory, 692–693. *See also* Reconsolidation
memory and
emotional effects on hippocampus, 729–730
sleep deprivation, 730–731

Constructive memory, 751–759
core network in, 754–756, 757–759
future event simulation, 752–754, 757–759
imagining future events, 752–757
past versus future events, 756–757
Consummatory responses, 914
Context
amygdala processing and, 935–942
context conditioning, 939–940
experimental context, 941–942
facial expression, 935–939
fMRI activation to specific stimuli, 936–937, 941–942
in cognitive neuroscience, 1259
in decision making, 1019–1020
in intention understanding, 635–636
in qualia space, 1212–1213, 1215
Contextual control, 1021
Contour integration, 130–131
Contour orientation, in discriminative learning, 133–134, 135, 136
Control. *See also* Cognitive (executive) control
interference, 1006–1007, 1013–1014
in semantic unification framework, 819
similarity with inference, 613–614
Convergent evolution, 876–877
Cooperativity, in memory formation, 113–114
Coordinate response measure (CRM), 349
Coordinate transformation, 541–543
Cornu ammonis 3 (CA3), 111
Corona radiata fibers, neuronal circuitry of frontal lobe, 36
Corpus callosum (CC), neuronal circuitry of frontal lobe, 31, 37, 39–40
Correlation
cortical, 316–318
grandmother cells, 316–318
temporal correlations in population response, 426–427
Cortical afferents
monkey entorhinal cortex, 667
monkey parahippocampal cortex, 665–666
monkey perirhinal cortex, 662–663
rodent entorhinal cortex, 667–669
rodent perirhinal cortex, 665, 666–667
Cortical architecture
cross-species comparisons of neocortical development, 7, 9–16, 19
designation of permanent cortical layers (I-VI), 31
future perspectives, 1247–1253
comprehensive theory of brain function, 1249–1250
modular hypothesis, 1247–1249
other complex systems, 1250–1251
model for modal texture filling-in, 442–447
stages of development, neuronal circuitry of frontal lobe, 29–44
of visual cortex, 130
Cortical column unit (CCU), 442
Cortical magnification factor (CMF), primate cortex, 422–423
Cortical mapping. *See also* Retinotopic projections/maps
human brain specialization, 53–54
motor cortex, 142–146
neuronal circuitry of frontal lobe, 30–31
protomap hypothesis, 18–19, 30–31
statistical connectivity and, 413–414
Cortical plate (CP)
cross-species comparisons of cortical development, 11, 13, 16–18, 20, 21
neuronal circuitry of frontal lobe, 29–32, 33–34, 35, 36, 43
Cortical projection neurons
cross-species comparisons of neocortical development, 8, 22–23
neuronal circuitry of frontal lobe, 33
Cortical thickness, 73–74
Cortico-cortical fibers (CC)
cross-species comparisons of cortical development, 13
neuronal circuitry of frontal lobe, 29, 32, 37, 39–40, 43–44
Corticogenesis
designation of permanent cortical layers (I-VI), 31
neuronal circuitry of human prefrontal cortex, 29–44
stages of
cross-species comparisons, 12, 13, 16
cytoarchitectonic layers, 31
early childhood, 30, 31, 41–44
early fetal period, 29–32
early preterm period, 30, 31, 33–36
embryonic period, 30, 31
human, 20–22, 29–44
infancy, 30, 31, 39–41
late fetal period, 30, 31, 32–33
late preterm period, 30, 31, 36–37
list of stages, 30
midfetal period, 30, 31, 32–33
neonatal period, 30, 37–39, 42
neural circuitry of prefrontal cortex, 29–44
Corticopontine, comparative neuroimaging, 57
Corticosterone (CORT), in infant fear learning, 893–896
Craik O'Brian Cornsweet (COC) stimulus, 438–441
Craniotopic maps, 519–521
CREB phosphorylation (pCREB), in attachment learning, 891, 893
Critical (sensitive) periods
for face recognition, 472, 479
for language development, 841, 850
for visual plasticity, 100, 130
Crowding task, video games and, 157, 158
Crying, in pathological laughter and crying (PLC), 1184
Crystallized reasoning, 78
Cue-induced amnesia, 693–694
Cue reliability, in multisensory integration, 506–507
Cultural factors, in processing of emotional information, 931–932
CYCLE, 809
Cytoskeleton
cross-species comparisons of cortical development, 14
neuroplasticity and, 146–147

D

Damasio, Antonio, 1183–1184
Darwin, Charles, 49–52, 873, 1242
Deafness
American Sign Language (ASL), 169–170
attention and, 170
motor detection and, 166, 167
Decision making
defined, 1019
emotion regulation in, 966–968, 1093–1100, 1187–1188
in neuroeconomics, 1075
computational basis, 1079–1081
emotion and, 1098–1100
initial investigations, 1098–1100
multiple behavior controllers, 1078–1079
neurobiological basis, 1079–1083, 1095–1100
simple binary stimulus choice, 1076–1077, 1079–1083
valuation process, 1085–1091
prefrontal executive function in, 1019–1027
action selection, 1021
cascade model, 1021–1023
hierarchical control, 1021
medial frontal cortex, 1023–1027
motivational control, 1023–1027
Declarative memory
awareness and, 686–687
emotional modulation of, 725–733
consolidation processes, 729–731
during encoding, 725–729
future directions, 733
during retrieval, 731–733
medial temporal lobe function, 675, 679–683, 686–687
recognition memory, 680–683
sleep deprivation and, 730
Declarative/procedural hypothesis, in morphological processing, 780–782
Decoding
dynamic state of posterior parietal cortex (PPC), 607–609
neural populations in primate cortex
different cortical areas, 430
different tasks, 430–431
mechanisms, 427–429
response, 430
in reading process, 797
Decorrelation, spatial, 427
Deep-brain electrical stimulation (DBS), following severe brain injury, 1131–1134
Deep homology
conservation of developmental pathways, 876–877
defined, 873–874
demonstration of, 879

discovery of, 877–878
evolutionary innovation in cognition and language, 877–878
Default node network, 1068, 1069
Defense conditioning, 907
Dementia
basal ganglia functions and, 568
fluid reasoning in, 79
frontotemporal dementia (FTD), 79
semantic, 1008, 1047, 1053–1061
Dendrites, neuronal circuitry of frontal lobe, 36–37, 40–41
Depictive theories, 1243
Depression, 147
Depth perception
neuronal versus behavioral sensitivity, 490
three-dimensional structural coding hypothesis, 460–463
visual area MT in, 483–496
binocular disparity processing, 483, 484–489
coarse and fine discrimination, 489–493
motion parallax, 484, 493–495
Depth sign discrimination index (DSDI), 495
Developmental biology
critical (sensitive) periods, 100, 130, 472, 479, 841, 850
defined, 876
marriage with evolutionary biology, 876
neural epigenesis, 876
Developmental stages
cross-species comparisons, 12, 13, 16
early childhood, 30, 31, 41–44
early fetal period, 29–32, 42
early preterm period, 30, 31, 33–36, 42, 43
infancy, 30, 31, 39–41, 470–473
language acquisition
neural basis, 837–851
visual processing of printed words, 791–794
late fetal period, 30, 31, 32–33, 42
late preterm period, 30, 31, 36–37, 42, 43
list of, 30
midfetal period, 30, 31, 32–33, 34
neonatal period, 30, 37–39, 42
neuronal circuitry of prefrontal lobe, 29–44
for visual plasticity
critical periods, 100, 130
early development, 129–130
late maturation, 130–131
Developmental verbal dyspraxia, 860–863
Diabetes, 858
Dieting, 957
Different band noise, 350
Different band sentence, 350
Differentiation level, 52–53
Diffusion-tensor and diffusion spectrum imaging (DTI/DSI), dense perisylvian white matter connectivity, 260–262, 263
Diffusion-tensor imaging (DTI)
of changes in interregional connectivity, 74
in comparative neuroimaging, 55–58
white matter structural changes following brain injury, 1127–1128
Direct realism, 847
Disconnection syndrome, 263–264, 269
Discourse models, in semantic unification framework, 822–824
Discriminative learning, 133–134
Disproportionate inversion effect, in face recognition, 468
Dissociation
consciousness and, 1142–1143
double dissociation in processing aphasia, 767–768, 780, 781
neuroanatomical dissociation hypothesis, 783–784
nonconscious processing of emotion, 1185–1188
and true function of consciousness, 1197–1199
Distinctiveness effects, in face recognition, 468
Distraction, 962–964, 969
anticipation in, 962
interference in, 962–964
Distributed algorithmic mechanism design (DAMD), 1251, 1252
Distributed attention, focused attention versus, 199
Domain-specific neural circuitry, 1035–1041
activity in, 1038–1040
property information in, 1035–1038
Dopamine
in basal ganglia function, 567, 571–572
in valuation process, 1087–1089
video games and, 161
Dorsal anterior cingulate (dACC), in cognitive control development, 77, 78, 254–255
Dorsal frontoparietal attention network. *See* Frontoparietal attention network, dorsal
Dorsal lateral geniculate (dLGN), 67–71
formation of eye-specific inputs, 68–70, 94
Dorsal visual pathway, 794–799
neuroplasticity of, 166–167, 168
reading process and
interfacing with verbal system, 798–799
orientation of attention, 795–796
parts of words, 797–798
pathology, 795–796
serial decoding, 797–798
single word selection, 796–797
timing of attention and, 193–194
visual awareness and, 1159–1160
Dorsolateral prefrontal cortex (DLPFC)
conflict-control model and, 252
development of, 74–76, 77, 79, 80, 705, 706, 707
in emotion regulation, 963, 966–967
memory and, 705, 706, 707, 719, 932
moral judgment and, 988, 992–994
in processing of emotional information, 932
in simple binary stimulus-choice paradigm, 1081–1083
Double-bouquet interneurons
cross-species comparisons of cortical development, 22
neuronal circuitry of frontal lobe, 41
Double Cortex, 14
Dragon software, 332–333
Driver, in relational reasoning, 1008
Dual-process morality, 991–993
Dual-task paradigm
attention versus consciousness, 1143
blood oxygen level dependent (BOLD) response, 1145
Dual task performance, as cognitive control function, 76
Dualism
conceptual, 1115
defined, 1114
explanatory gap in consciousness and, 1114–1115
Dynamic belief networks, 614–621
Dynamic causal model (DCM), effective connectivity of fMRI data, 375–378
Dyslexia, 166, 167, 168. *See also* Reading process
attentional, 797
central, 789
developmental, 857–859, 863–866
neglect, 795–796
peripheral, 789
spatial, 798
Dyspraxia, 860–863
Dystonia, 565, 572

E

Early childhood stage. *See* Childhood
Early infancy stage
defined, 30
neuronal circuitry of frontal lobe, 30, 31, 39–41
functional organization, 39–41
neurogenetic events, 39
structural organization, 39
Early language acquisition, 837–851
infant lexicon, 843–844
neuroscience-based measures, 837–839
phonetic learning, 839–841
second-language learning, 842, 848–849
sentence processing, 844–845
word learning, 842–843
Early preterm stage
defined, 30
laminar shifts in acetylcholinesterase (AChE), 34
neuronal circuitry of frontal lobe, 30, 31, 33–36, 42, 43
functional organization, 35–36
neurogenetic events, 34–35
structural organization, 33–34
Early Reading Intervention (ERI), 175
Ecological self, 954, 955
EEG/MEG studies
brain measurement of auditory object analysis, 371, 376
physiological mechanisms in visual object selection, 229–230
of semantic unification, 822
Effective information, 1204
Efferent associative pathways, neuronal circuitry of frontal lobe, 37, 43
Egly paradigm, 242

18F-fluorodeoxyglucose (FDG)-PET technique, 57
Elaboration, effects of emotion on, 728
Electroconvulsive shock (ECS), in test of consolidation theory, 693
Electroencephalography (EEG). *See also* EEG/MEG studies
  early language acquisition, 837–838, 839, 849–850
  empathy and, 982
  neural processing of emotional information, 925–926
  spontaneous motor initiation, 1191
Electromyelography (EMG)
  empathy and, 974–975, 981
  motor cognition in autism, 636–638
  muscle synergies, 544–546
Elliot Smith, Grafton, 51
Embodied cognition, 1032–1035
Embryonic stage
  defined, 30, 31
  formation of visual pathway, 91–92
Emotion. *See also* Emotion regulation; Empathy; Fear learning/conditioning; Moral judgment; Social cognition
  amygdala in processing. *See* Amygdala
  attention and, 190, 925–932
    automaticity, 929–931
    behavioral effects on perception, 926–928
    cultural factors, 931–932
    current behavior goals, 932
    neural circuits underlying attention, 928–929
    neural processing in perception, 925–926
    personality factors, 931–932
  consciousness and, 1181–1188
    conscious experience of emotion, 1181–1185
    nonconscious processing of emotion, 1185–1188
  context effects and amygdala, 935–942
    context conditioning, 939–940
    experimental context, 941–942
    facial expressions, 935–939, 941–942
    fMRI activation to specific stimuli, 936–937, 941–942
  feelings versus, 1182. *See also* Feelings
  imaging genetics and, 945–949
    amygdala reactivity, 947–948
    conceptual basis of research, 945–946
    importance of, 946
    neuroimaging role, 945–946
    principles of imaging genetics, 946
  in modulation of declarative memory, 725–733
    consolidation processes, 729–731
    during encoding, 725–729
    future directions, 733
    during retrieval, 731–733
  object concepts in support of feeling, 1032–1035
  reaction and action, 905–919
    behavioral distinctions, 905–909
    neural basis of fear actions, 914–918
    neural basis of fear reactions, 909–914
  regulation of. *See* Emotion regulation
  social cognition and, 1184–1185. *See also* Social cognition
Emotion regulation, 961–970
  in decision making, 966–968, 1093–1100, 1187–1188
    amygdala, 1095–1097
    initial investigations, 1098–1100
    striatum, 1094–1095, 1097
  defined, 961
  distraction, 962–964, 969
  future directions, 969–970
  methodology for study of, 961–962
  neuroimaging research on, 968–970, 1099
  reappraisal, 964–966, 969
  self-regulation, 956–957
Emotional attention, 925
Emotional attention sets, 932
Emotional contagion, 974
Emotional memory, 700, 725
Empathy, 973–982. *See also* Social cognition
  defined, 973–974
  healthy versus clinical populations, 978–979
  insular involvement, 977–978
  modulation of, 979–981
  moral judgment and, 988–989
  prosocial behavior and, 982
  related concepts, 973–974
  shared networks hypothesis, 974–977
Encoding
  dynamic state of posterior parietal cortex (PPC), 604–607
  emotional influences on, 725–729
  of emotional information, 728
  neural populations in primate cortex, 419–432
    mean response, 422–425
    response variability, 425–427
    theoretical framework, 420–421
    in vivo measurement, 421–422
  in olfaction
    odor, 326–327
    from percept to molecule, 330
    spatial, 327–329
    temporal, 329–330
  temporal
    odor discrimination, 329–330
    perceptual learning, 135–136
  variable nature of episodic memory, 743–744
Encyclopedic knowledge, in conceptual representations, 1032
End-stopped neurons, 314
Endogenous attention, 925, 932
Endpoint coordinates, 541–542
Energetic masking, auditory, 344–345, 349, 351
Entanglement
  consciousness and, 1141
  in qualia space, 1213
Entorhinal cortex, comparative anatomy of, 661, 662, 667–671
Entropy, in integrated information theory (IIT), 1202, 1204
Ephrin-A receptors (EphAs), 70, 92–94, 97, 98
Ephrin-B receptors (EphBs), 92–94
Episodic control, 1021
Episodic memory
  as constructive, 751–759
    future event simulation, 752–754, 757–759
    imagining future events, 752–757
    past versus future events, 756–757
  distribution throughout cerebral cortex, 743–744
  encoding/retrieval studies, 712, 731–733
  individual differences and, 743–747
    neural basis of episodic retrieval, 743
    sources of variability, 744–747
    variable nature of brain regions, 743–744
  parahippocampal region and, 670
  sleep deprivation and, 730
Epithelium, olfactory, 324–325, 327–328
ERP. *See* Event-related potentials (ERPs)
Error/feedback processing, as cognitive control function, 76
Error signals, in learning and learning transfer, 160
Escape from fear hypothesis (EFF), 907–909, 1096
  amygdala contributions to conditioned reinforcement, 914–917, 1095–1097
  conditioned motivation, 909, 915–917
  described, 907–909
Essential genes, cross-species comparisons of neocortical development, 7–8
Estes, William, 740–741
Estimation. *See* Optimal estimation
Ethics, of cognitive neuroscience, 1260
Event knowledge, in semantic unification framework, 822–824
Event-related magnetic fields (ERMFs), in analysis of attention development
  feature-based attention, 238–240
  object-based attention, 240–246
  spatial attention, 235–238
Event-related potentials (ERPs)
  in analysis of attention development, 170, 176, 186, 193, 196–197
    biased competition theory, 199
    conflict and attentional control, 254–255
    feature-based attention, 238–240
    object-based attention, 240–246
    spatial attention, 235–238
  in auditory development, 172–173, 175
  basal ganglia function and, 577–578
  early language acquisition, 169–170, 837–838, 839–849
  in emotion regulation, 964
  face recognition and, 469, 472–473, 474–477
  integration of conflict processing and attention, 253–255
  in language development, 169–170
  neural processing of emotional information, 929, 930

in semantic unification, 820–822, 823, 824–827
spontaneous motor initiation, 1192
in vision, 166, 167
Evolutionary biology, 49–53. *See also* Genetics; Genomics
cerebral cortex in, 314–316
consciousness in, 1138–1139
Darwin and, 49–52
fear learning in, 892
human brain specialization, 53–61
human brain versus other animals, 49–52
language in, 873–881
deep homology, 876–878
evo-devo perspectives, 876
human cognitive evolution, 874–876
mechanisms underlying speech and vocal learning, 878–879
semantics, 875, 879–880
syntax, 880
model-animal versus comparative approaches, 52–53
time periods, 51
Evolutionary developmental biology (evo-devo), 876
Excitatory neurons, cross-species comparisons of cortical development, 22–23
Excitatory projection cells, cross-species comparisons of neocortical development, 7, 22–23
Exclusion
conscious, 1194–1195
noise, 171, 173
Executive control. *See* Cognitive (executive) control
Exogenous attention, 925, 932
Experience
in consciousness, 1137–1138
giving shape to, in qualia space, 1212–1213
in isolation, face recognition and, 469
Experience-dependent learning, 154–155, 161, 167
Experimental context
instructions, impact of, 941–942
neutral position, 941
static versus dynamic, 941
Explanatory gap in consciousness, 1113–1115
described, 1113–1114
dualism and, 1114–1115
Expressed knowledge, in conceptual representations, 1032
Expression studies, nature of, 857
Extracellular signal-regulated kinase-mitogen-activated protein kinase (ERK-MRPK), 696
Extrastriate body areas (EBA), neural processing of emotional information, 926
Extreme capsule (EmC), 261
Eye movements
in attentional selection, 290
eyeblink conditioning studies, 687, 688
in feature-based selection, 226–227
memory and changes in, 687
saccades, 511–521
forward model for eye position, 600
localization error and, 514–516
in motor learning and control, 591–592
reafference cancellation in PPC, 600–601
saccadic compression, 515
saccadic suppression, 511–514, 1154
transsaccadic integration and craniotopic maps, 519–521
updating of internal spatial maps, 514–519
visual stability and, 511
transitive inference studies, 687, 688
value-related signals, 1086

F

F5 neurons, in grasping behavior of monkeys, 626–632
Face recognition
in adulthood, 467–469
core behavioral properties in humans, 467–468
electrophysiological signatures in humans, 469
neurophysiology in monkeys, 468–469
roles of experience and genetics, 469–470
behavioral measures, 467–470, 473–474, 477–478
development
for behavioral and neural measures, 477–478
in four-year-olds to adults, 473–478
in infancy, 470–473
grandmother cells in, 312–314
in infancy
critical (sensitive) period and, 472
face individuation, 470–471
perceptual narrowing, 471–472
neural measures, 468–469, 474–478
Facial expression
empathy and, 974–975
in study of human amygdala, 935–939
experimental context, 941, 942
fear, 935–936
surprise, 936–939
Faculty of language in a broad sense (FLB), 873, 875
Faculty of language in a narrow sense (FLN), 875
False memories, 700
Familiarity, medial temporal lobe function, 680–683
Fear learning/conditioning, 698–699. *See also* Pavlovian conditioning
avoidance conditioning, 116, 907–909
detection of threat, 957–958
escape from fear (EFF) hypothesis
conditioned motivation, 909, 915–917
described, 907–909, 1096
escape from fear hypothesis (EFF), amygdala contributions to conditioned reinforcement, 914–917, 1095–1097
facial expression and amygdala response, 935–936, 941, 942
infant, 891–896
amygdala in, 891–896
corticosterone in, 893–896
maternal separation/deprivation, 895
neonatal handling, 895–896
rearing environment alteration, 895
learned fear response, 906
neural basis for fear actions, 914–918
neural basis for fear reactions, 891–896, 909–914, 930, 1095–1097
research methodology, 906–907
Feature-based selection, 197–198
attentional-biasing signals, 208–210
eye movements in, 226–227
loss in Balint's syndrome, 271–272
selective neuronal synchronization and, 297–298
in spatiotemporal analysis of visual attention, 238–240
Feature binding, attention and, 197–198, 274–276
Feature integration theory, attention and, 197–198, 208–210, 226–227, 276–277
Feedback
within and across populations, 431
basal ganglia function and, 575–577
as cognitive control function, 76
forward versus feedback projections, 1145–1146
in learning and learning transfer, 160–161
recurrent feedforward/feedback loops, 438
visual attention and awareness, 1165–1175
anatomical observations, 1165–1166
binocular rivalry role in, 1170–1176
physiological observations, 1166–1168
role of feedback in attention, 1168–1170
visual masking role, 1170–1176
Feedback control
in motor learning and control, 588, 592
in optimal control theory, 614–622
Feelings. *See also* Consciousness
emotions versus, 1182. *See also* Emotion
meaning of, 1138
as private, 1138
Ferrets
auditory cortex, rewiring vision into, 97–98
visual cortex, formation of eye-specific inputs to dLGN, 69, 94–95
Fetal stage
early, defined, 30
late, defined, 30
neuronal circuitry of frontal lobe, 29–32, 30, 31, 32–33, 35, 42
functional organization, 32, 33
neurogenetic events, 31
neurogenic events, 32–33
structural organization, 29–31, 32
Field excitatory postsynaptic potential (fEPSP), 109–112, 118–121
Fight-or-flight responses, 155
Filling-in. *See* Perceptual filling-in

Filtering
attention and, 171
auditory, 345–348
Fine depth perception, 489–493
Firing rate, 290–291
Fish, vision, activity-dependent refinement of visual maps, 98–99
Fixed action patterns, 906
Flanker interference, 797
Flash suppression, 1142, 1144, 1145, 1154
Fluid reasoning
analogical, 78, 79
assessing reasoning ability, 78–80
brain change and improvements in, 80
development of, 73, 77–81
neural correlates of, 79–80
timing of development of, 78
Focal hand dystonia, 147
Focused attention, 192–193
distributed attention versus, 199
Footbridge dilemma, 991–992
Formal operations, 78
Fossil record, 52
FOXP2 (forkhead box P2), 59, 861–866, 878, 879
Fragile X syndrome, 147, 166
Free will. *See also* Consciousness
neurobiology of, 1138
Freud, Sigmund, 1033
Frogs, modularity in spinal motor system, 546–547
Frontal eye field (FEF)
anatomy of feedback in primary visual cortex, 1165–1166
conflict and cognitive control, 255–257
and dorsal frontoparietal attention network, 219–227
eye movements and, 517, 518, 600
virus tracing studies, 556–557
Frontal lobe circuitry, early development of prefrontal cortex, stages of development, 29–44
Frontoparietal attention network, 219–227
dorsal
anticipatory signals, 219–223
causality of top-down biases, 223
coding locus of attention, 224–226
conflict and cognitive control, 251–257
conflict-control model, 252
eye movements, 226–227
feature-based selection, 226–227
functional connectivity, 219–223
topographic maps/projections, 223–226
establishing goals
reward/value signals and limbic system, 227–228
task sets, 228–229
nonemotional, task-relevant information and, 727
physiological mechanisms for selection of visual objects, 229–230
prefrontal-cingular circuits, 228–229
ventral
neuroplasticity of, 166–167
relation to dorsal attention network, 219–222
timing of attention and, 193–194
visual awareness and, 1157–1158
working memory, 228–229
Frontotemporal lobal degeneration (FTLD), in relational reasoning, 1008–1015
Frozen addicts, 1141
Functional connectivity MRI (fcMRI)
comparative, 57–58
dorsal and ventral attention systems, 219–221
Functionalism, 1111
Functional knockout, 858
Functional magnetic resonance imaging (fMRI)
absolute and relative disparity selectivity, 486–489
action understanding, 647–648
age differences in brain activation, 82
attention and
conflict processing, 253–255
connectivity studies, 264
control of attention, 198
event-related potential (ERP), 193
feature-based attention, 209–210
middle temporal (MT) changes, 190
object-based attention, 208
object selection, 242–244
sequential/simultaneous paradigm, 212–213
space-based attention, 206–208
task sets and working memory, 228–229
temporal correlation of BOLD signal, 219
auditory object analysis, 371–379
analysis of categorical processing, 374–375
effective connectivity analysis, 375–379
multivariate analysis, 374
univariate analysis, 372–374
basal ganglia function and, 576
blood oxygen level dependent (BOLD) response, 313
auditory object analysis, 371–376
degree of similarity in patterns of brain activity, 742–743
genetic basis of emotional variability, 945–949
individual differences and, 742
network-brain links to visualize spiking, 435, 443, 444, 446
neuronal basis of perceptual illusions, 1144–1145
semantic unification and, 827–832
syntactic processing, 810–813
visuospatial working memory, 74
brain function and, 1067, 1068, 1070
comparative, 57–58
contour detection, 135
detecting awareness following brain injury, 1125–1127
emotion regulation, 963–968, 970, 1099
empathy and, 975–977, 981
episodic memory tasks, 739
eye movements and, 520–521
of face recognition in, 468–469, 477
fMRI mutation, 147
integration of conflict processing and attention, 253–255
interactions between control and memory, 712, 713
language processing
developmental verbal dyspraxia, 861
early language acquisition, 837–838, 839, 850
morphological processing, 782, 784
syntactic, 810–813
in methodology of cognitive neuroscience, 1255–1256, 1257–1258, 1260
mirror neural system studies, 635, 636
moral judgment and, 991, 992–993
neural processing of emotional information, 925–926, 927, 929, 930
object concepts and, 1040
perceptual filling-in, 442–449
phonological processing and, posterior language cortex in speech production, 770
posterior parietal cortex (PPC) studies, 599–600
processes in relational reasoning, 1011–1015
reading ability and, 175–176, 799
in semantic memory processing, 1062, 1063–1064
simulation of future events, 754–755
spiking activity in surface perception, 435, 443, 444, 446
spontaneous motor initiation, 1192, 1193
value-related signals, 1087–1088, 1091
visual attention and, 205–214
pupillary response, 190
selection among multiple competing objects, 210–214
units of selection, 206–210
visual plasticity and, 166
Functional maps, in primary visual cortex, 409–411
Functional organization of brain, neuronal circuitry of frontal lobe, 32–42
Functional specialization, in Visual Word Form Area (VWFA), 794
Fusiform body areas (FBA), neural processing of emotional information, 926
Fusiform face area (FFA)
conflict-control model, 252
cortical loci of face identity processing, 468–469
in face recognition, 469, 474–479
neural processing of emotional information, 926
visual rivalry and, 1156

G

GABAergic cortical interneurons
activity-dependent refinement of visual maps, 99
cross-species comparisons of neocortical development, 8–9, 12, 22, 23
neuronal circuitry of frontal lobe, 35–36, 41, 42–43
Gabor function, 316
Gabor patches, 212, 243, 425, 427
Galanin, in comparative histology, 55

Gambling, 968, 1098–1100, 1187–1188
Gamma-band synchronization, 292–298
Ganglionic eminence (GE), neuronal circuitry of frontal lobe, 31
Gaussian priors, 615, 616, 617
Gazzaniga, Michael, 1221, 1233, 1255
Gender, in lesion studies of language function, 741
Gene transcription, in ocular dominance plasticity, 100–102
Generalist genes, 863
Generalization
  in auditory training, 355–358, 360
  in color vision, 396–406
  overgeneralization, 1048, 1051–1052, 1059
Genetics. *See also* Evolutionary biology; Genomics
  of emotional variability, 945–949
    amygdala reactivity, 947–948
    conceptual basis of research, 945–946
    importance of, 946
    neuroimaging role, 945–946
    principles of imaging genetics, 946
  inherited color vision deficiencies, 388–392
  of language impairment, 855–856
    approaches to study of, 856–857
    developmental dyslexia, 857–859
    developmental verbal dyspraxia, 860–863
    evidence, 855–857
    perspectives, 863–866
    specific language impairment, 859–860
    speech sound disorder, 860
  in methodology of cognitive neuroscience, 1257
Genomics
  comparative, 58–61
  gene sequencing, 58–61
  phenotype changes, 59–61
Genotypes, 876
Geometric cues, 483
Gestalt psychology, 212–213, 317
Gibbs distribution, 615
Gibbs sampling, 622
Glial acidic fibrillary protein (GAFP), cross-species comparisons of cortical development, 9, 22
Global workspace approach to consciousness
  described, 1111–1112
  explanatory gap and, 1114
  machine consciousness and, 1119–1120
  sensory activation in, 1112, 1146, 1160
Globus pallidus (GPi)
  basal ganglia function, 569–571
  central thalamic deep-brain stimulation, 1131
  virus tracing studies, 553–556
Glomerular modules, 328
Glossogeny, 875
Glutamate cycling, 1070
Glutamate receptors, 112–113
Glutamergic neurons, 35, 42
Glutamergic thalamic afferents, neuronal circuitry of frontal lobe, 36–37, 43–44
Glycoprotein, 94
Gnostic neurons (Konorski), 309–312. *See also* Grandmother cells (visual cortex)
Goal-directed choice, 1075–1083. *See also* Goals; Neuroeconomics
  computational basis, 1079–1081
  multiple behavior controllers, 1078–1079
  neurobiological basis, 1079–1083, 1095–1100
  simple binary stimulus choice, 1076–1077, 1079–1083
Goal representation. *See also* Goals
  in bimanual coordination, 645–647
  hierarchical nature of, 642–644
  on-line control of grasp, 644–645
Goals. *See also* Goal-directed choice; Goal representation
  frontoparietal network and, 227–229, 626–627
  in learning and learning transfer, 160–161
  premotor cortex and, 1037–1039
Grammar
  agrammatic aphasia, 808
  artificial grammar learning, 880
  components of, 1235–1237
Grammatical categories
  in morphological processing, 782–785
  neuroanatomical dissociation hypothesis, 783–784
Grandmother cells (visual cortex), 309–319
  correlation, 316–318
  cortical function, 314–316
  defined, 310
  history, 309–314
  invariance, 316–318
  symmetry, 316–318
Granger causality analysis, 223
Grasping behavior
  goal representation and, 644–645
  mirror neural system and, 626–638
Gray matter. *See also* White matter (WM)
  in cognitive control development, 76
  in fluid reasoning development, 80
  and plasticity of human neurocognition, 165
  thickness of, 73–74
Great Chain of Being, 49–50
Grid cells, 121
Gross, Charles, 309–310, 312–313, 318
GY (patient), nonconscious processing of emotion, 1186, 1194–1195, 1199

H

Habits
  goal-directed choice and, 1078, 1083
  medial temporal lobe function and, 679–680
  nature of, 906
Haddad, Rafi, 334–335
Haeckel, Ernst, 50
Hallucinations, 1156
Hamilton-Jacobi-Bellman (HJB) equation, 549
Hand-eye coordination
  in athletics, 155, 156–157
  video games and, 157–158
Hand motor function
  focal hand dystonia, 147
  recovery after stroke, 148–149
  specificity of learning, 154
Hawthorne effect, 155
Heading discrimination task, 500–502, 503–505
Heart disease, 858
Hebb, Donald, 109, 129
Hebbian rule of synaptic plasticity, 99, 100–101, 129
Hebrew, classical, functional morphemes in, 777
Helmholtz, Hermann von, 132, 137, 305, 321, 514, 530, 621
Hemianopia, 792, 1159
Herrick, C. J., 316
Heschl's gyrus (HG), 372–373, 377–378, 1237
Hierarchical control
  in decision making, 1019, 1021
  forms of, 641–642
  memory and, 707–709
  in representation of action, 641–650
    action semantics, 648–650
    action understanding, 647–648
    anatomical versus representational hierarchy, 642–644
    bimanual coordination, 645–647
    goal representation, 644–647
    grasp control, 644–645
    hierarchy of serial behavior, 641–642
    historical perspective, 641–642
  reverse hierarchy theory (Hochstein and Ahissar), 194
  in surface perception, 437–438
Hierarchical Linear Growth Curve modeling, 840
High-frequency stimulation (HFS), 111, 112, 117
Higher order consciousness, described, 1116
Higher order thought (HOT) approach to consciousness
  consciousness of and, 1115–1116, 1182
  described, 1111
  explanatory gap and, 1114
  machine consciousness and, 1119–1120
  problems in, 1115–1118
  self in, 1118
  sensory activation in, 1112
Hillyard principle, 170
Hinckley, John, Jr., 1260
Hippocampus
  amygdala modulation of consolidation in, 729–730
  attractor networks, 121–123
  basal ganglia function and, 575–576
  lesion studies of medial temporal lobe function, 675–678, 681–682, 685–686
  long-term memory and, 229
  long-term potentiation (LTP) studies, 109–111, 116, 119–121
  response to simulation of future events, 756, 759
Histamine, 1140

Histology, human brain specialization, 54–55
HM (amnesia patient), 656, 659, 675
Hodological studies, mirror neural system in monkeys, 626
Holistic/configural processing, in face recognition, 468
Holmes, Oliver Wendell, Jr., 993
Homeostatic afferent system, 1183–1184
Homeostatic processes, in ocular dominance plasticity, 101
Hominoids
  brain specialization, 53–61
    comparative genomics, 58–61
    comparative histology, 54–55
    comparative neuroimaging, 55–58
  types of, 53
Homunculus, 1152–1154, 1248–1249
HOX genes, 876, 878, 879
Hubel, David, 309, 311–312, 314, 409
Humphrey Field Analyzer (HFA-II-I), 1119
Huntington's disease, 561
  basal ganglia functions and, 565, 568, 573–574, 577
Huxley, Thomas H., 52
Hyperdirect pathway, in basal ganglia function, 568–569
Hypocretin, 1140
Hypothalmic-pituitary-adrenal (HPA) responsiveness, 893, 895–896

I

Illusions
  Kanisza illusion, 213
  neuronal basis of, 1143–1145
Illusory correlation, 1051–1052
Imagery debate, 1241–1245
  general lessons from, 1244–1245
  relevance of the brain, 1243–1244
  theory of cognition, 1241–1243
Imaging genetics, 945–949
  amygdala reactivity and human emotion, 947–948
  basic principles, 946
  conceptual basis, 945–946
  importance of, 946
  neuroimaging in, 946
Implicit memory, in perceptual learning, 132
Impulse-control disorders, basal ganglia function and, 572–573
Incest, moral judgment and, 990
Indirect pathway, in basal ganglia function, 568–569
Individual differences, 733
  dangers of averaging across subjects, 739–741
  episodic memory and, 743–747
    sources of variability, 744–747
    variable nature of episodic memory, 743–744
  genetic basis of emotional variability, 945–949
  moral judgment and, 992
  neuroimaging and, 741–742
  in processing of emotional information, 931–932
  similarity between two patterns of brain activity, 742–743
Infancy. *See also* Developmental stages
  attachment learning, 889–891, 894
  differential development and, 1048
  early language acquisition, 837–851
    bilingual infants, 848–849
    infant lexicon, 843–844
    neuroscience-based measures, 837–839
    phonetic learning, 839–841
    second-language learning, 842, 848–849
    word learning, 842–843
  face recognition in, 470–473
  fear learning, 891–896
  phenomenal consciousness during, 1117–1118
Inference
  in Bayesian statistics, 159, 614–622
  and higher order of thought (HOT) approach to consciousness, 1116
  similarity with control, 613–622
  transitive inference studies of eye movements, 687, 688
Inferior frontal cortex (IFC), in semantic unification network, 827–832
Inferior frontal gyrus (IFG)
  action understanding, 647–648
  language processing and, 260, 811–813
  mirror neural system and, 633–634
  modularity of language and, 1236–1238
  in specific language impairment (SLI), 859
Inferior occipitofrontal fasciculus (IOF), dense perisylvian white matter connectivity, 261
Inferior parietal lobule (IPL)
  mirror neural system and, 626–638
  perisylvian network for spatial orienting, 259–260, 263
Inferotemporal (IT) cortex
  in depth perception, 489
  virus tracing studies, 558–559
Inflections, 777
Information, in integrated information theory (IIT), 1202–1204
Information detection, effect of emotion on, 726–727
Information theory, 1146–1147
Informational masking, auditory, 345–351
  simultaneous multitone maskers, 345–348
  speech recognition and, 348–351
Informational relationships, in qualia space, 1212, 1213
Inhibition
  of attention, 193
  conscious, 1194–1195
Inhibitory avoidance training, 116
Inhibitory interneurons, cross-species comparisons of neocortical development, 7
Input specificity, in memory formation, 113, 114
Instinct to learn, 873
Integrated information theory (IIT), 1111–1112, 1147, 1202–1216
  complexes in, 1205, 1207
  down-set of red, 1215
  information in, 1202–1204
  integration in, 1203, 1204–1205, 1207
  measuring integrated information, 1203–1205
  neurobiological observations, 1205–1209
  up-set of nonred, 1215
Integration control, in relational reasoning, 1006–1007
Intelligence quotient (IQ)
  effect of music lessons on, 156
  experience-dependent learning and, 154–155
  in lesion studies of language function, 741
  measuring, 154–155, 156
  Mozart effect, 154–155
  music lessons and, 156
  socioeconomic factors in, 80, 171
  timing of cortical maturation in frontal regions, 80
Intelligence, integrated information theory of, 1112
Intention
  in intention understanding, 632–633, 635–636
  movement anticipation and, 599–601
Interaural level differences (ILDs), 359–362
Interaural time differences (ITDs), 359–362
Interface hypothesis of attention, 213–214
Interface mechanisms, in phonological processing, 768
Interference, 962–964
  defined, 709
  interactions between control and memory, 710–713
  interference suppression/resolution, 76, 714–719
  reduced, 714–719
Interference control, 1006–1007, 1013–1014
Intermediate sensory representations, 619–621
Intermediate zone (IZ)
  cross-species comparisons of cortical development, 11, 13, 20, 21
  neuronal circuitry of frontal lobe, 31, 33
International Affective Picture Set (IAPS), 964, 966, 970
Internet, ad auctions and, 1250–1251
Interoceptive system, 1183–1184
Interpersonal Reactivity Index (IRI), 978–979
Interpolation
  defined, 436–437
  in surface perception, 439–442
Interpretation, of stimulus, 1038
Interstitial neurons, cross-species comparisons of cortical development, 10
Intracortical connections, neuronal circuitry of frontal lobe, 39
Intralaminar nuclei of the thalamus (ILN), 1140
Intraparietal sulcus (IPS)
  and dorsal frontoparietal attention network, 219–227
  neural correlates in PPC, 604
  neural processing of emotional information, 929

social cognition and, 955
in visuospatial working memory, 74
Intrinsic projections
cost of intrinsic activity, 1068–1070
monkey entorhinal cortex, 667
monkey parahippocampal cortex, 665–666
monkey perirhinal cortex, 663–665
organization of intrinsic activity, 1070–1071
rodent entorhinal cortex, 669
rodent perirhinal cortex, 665, 667
Introspection, 1141–1142
Iowa Gambling Task (IGT), 1098–1099, 1187–1188
Ipsilateral cortex, 94, 98–99, 123, 148
Isolating languages, functional morphemes in, 777
Isometrics
defined, 614
duality of Bayesian inference and optimal control, 614–615

J

Jackson, John Hughlings, 741
James, William, 129, 173–175, 954, 1031, 1182–1183
James-Lange theory of emotion, 977, 1182
Jaynes, E. T., 530
Jebsen Taylor Hand function test (JTT), 149
Joint coordinates, 542–543
Journal citation maps, 1222–1230
developing, 1222–1226
interpreting, 1226–1230

K

Kalman filter, 602–603, 608–609, 617, 621–622
Kalman gain, 590, 603
Kamada-Kawai algorithm, 1230
Kanisza illusion, 213
Kant, Immanuel, 269, 270
KCC transporter, 36
KE family (multigenerational pedigree), 860–863, 864–865
Khan, Rehan, 330
Konorski, Jerzy, 309–312
Konorski, John, 129
Kullback-Liebler (KL) divergence, 618, 1204
Kuypers/Jürgens hypothesis, 878

L

L cone mosaic
Bayesian statistics and, 401–404
described, 395
functional consequences of, 386–388
structural organization of, 386, 387
Laminar domain
bilaminar distribution of early synapses, 32
neocortical development
cross-species comparisons, 7, 18–19
shifts in acetylcholinesterase (AChE), 34
Lancet, Doron, 324
Language impairment, 166, 171, 259–260. *See also specific types*
genetics in, 855–856
approaches to study of, 856–857
developmental dyslexia, 857–859
developmental verbal dyspraxia, 860–863
evidence, 855–857
perspectives, 863–866
specific language impairment, 859–860
speech sound disorder, 860
in reading process, 792–798
Language processing. *See also* Language impairment
attention training for children, 173–174, 176
auditory processing and, 168–169
Chomsky's model, 806, 807–808, 837, 855
cognitive neuroscience of, 1235–1239
future directions, 1238–1239
modularity of language, 1235–1238
right hemisphere in language processing, 1238
consciousness versus, 1141–1142
early language acquisition, 837–851
in autism spectrum disorder (ASD), 845–847
bilingual infants, 848–849
infant lexicon, 843–844
mirror neurons and shared brain systems, 847–848
neuroscience-based measures, 837–839
phonetic learning, 839–841
second-language learning, 842, 848–849
sentence processing, 844–845
word learning, 842–843
in evolutionary biology, 873–881
deep homology, 873–874, 876–878
evo-devo perspectives, 876
human cognitive evolution, 874–876
mechanisms underlying speech and vocal learning, 878–879
semantics, 879–880
syntax, 880
genetics of language, 855–866
developmental dyslexia, 857–859
developmental verbal dyspraxia, 860–863
evidence, 855–857
perspectives, 863–866
specific language impairment, 859–860
speech sound disorder, 860
morphological processes, 777–785
grammatical categories, 782–785
morphological composition in the lexicon, 778
neural basis, 778–780
regular/irregular debate in, 780–782
perisylvian network for spatial orienting, 259–260
phonological, 767–774
auditory-motor integration networks, 771
defined, 767
left posterior planum temporale (area Spt) in, 770–771
mirror neurons in, 773
motor theories of perception in, 773
posterior language cortex in left hemisphere, 770
speech recognition, 769–773
spoken word recognition, 768–769
superior temporal sulcus (STS) in, 769
task dependence, 767–768
plasticity in development, 169–170
reading process, 789–800
dorsal visual pathway, 794–798
interfacing with verbal system, 798–799
ventral visual pathway, 789–794
semantic. *See* Semantic processing
signed language, 169–170
syntactic, 169, 805–814
lesion studies, 807–810
neuroimaging studies, 810–813
syntactic representations and processing, 805–807
syntax-semantics framework, 824–827
Wernicke-Geschwind model, 741
Late infancy stage
defined, 30
neuronal circuitry of frontal lobe, 30, 31, 39–41
functional organization, 39–41
neurogenetic events, 39
structural organization, 39
Late positive potentials (LPP), neural processing of emotional information, 926
Late preterm stage
defined, 30
neuronal circuitry of frontal lobe, 30, 31, 36–37, 42, 43, 44
functional organization, 37
neurogenetic events, 36–37
structural organization, 36
Lateral amygdala (LA), in fear reaction, 911–912, 914–915, 918–919
Lateral entorhinal area (LEA), comparative anatomy of, 667–669, 671
Lateral frontal cortex (LFC), in decision making, 1021–1023
Lateral geniculate nucleus (LGN)
baseline activity in spatial attention, 206–208
eye movements and, 513
feedback in, 1165–1166, 1168–1169, 1173
in formation of visual pathway, 92–95
grandmother cell research, 309, 315
physiology of feedback in, 1165–1166, 1168–1170
receptive fields in visual cortex, 410–413
in rewiring vision into auditory pathway, 95–98, 167
visual rivalry and, 1156
Lateral intraparietal cortex (LIP)
eye movements and, 516–518
in valuation process, 1089, 1090

Lateral prefrontal cortex, in emotion regulation, 964–965
Laughter, in pathological laughter and crying (PLC), 1184
Layer I pyramids, neuronal circuitry of frontal lobe, 31, 37, 44
Layer II pyramids, neuronal circuitry of frontal lobe, 37
Layer III pyramids, neuronal circuitry of frontal lobe, 29–32, 36–41, 44
Layer IIIC pyramids, neuronal circuitry of frontal lobe, 37, 38, 39, 41, 44
Layer IV pyramids, neuronal circuitry of frontal lobe, 31, 32, 36, 37, 41
Layer V pyramids
  in comparative histology, 54
  neuronal circuitry of frontal lobe, 29, 33, 37, 38, 39, 41, 44
Layer VI pyramids
  neuronal circuitry of frontal lobe, 31, 36–39, 92–98
  visual awareness and, 165–1168, 1155–1156, 1159, 1170, 1171, 1172–1176
  visual plasticity in postnatal development, 129–132, 136–137
Le Gros Clark, W. E., 51
Learned fear response, 906
Learning
  basal ganglia function and, 375–377
  experience-dependent, 154–155, 161, 167
  fear conditioning, 116, 118, 698–699
  infant
    attachment, 889–891, 894–896
    fear, 891–896
  memory versus, 655
  motor. *See* Motor processing
  perception versus, 675
  rewards in, 161, 588, 592–593
  in simple binary stimulus-choice paradigm, 1081, 1083
Learning transfer, determinants of, 158–161
LeDoux, Joseph, 696
Left visual field (LVF)
  in attending parts of words, 797–798
  in early visual processing of printed words, 791–792
  in semantic processing, 820
  in serial decoding, 797–798
Lesion studies
  of action semantics, 649–650
  of aphasia, 1236
  of basal ganglion function, 571, 575–576, 594
  in conditioned fear reactions, 914–917, 1096–1097
  of decision making, 1094
  of emotion regulation, 968–969, 970
  implicit perception and, 195–196
  of language function, 741
  of medial temporal lobe function, 675–686
  in methodology of cognitive neuroscience, 1256
  of morphological processing, 779, 782
  paradoxical effect of brain, 142–143
  parahippocampal function in rats, 660, 670
  phonological processing and, posterior language cortex in speech production, 770
  plasticity as opportunity for intervention, 148
  posterior parietal cortex (PPC) function, 601, 603–604
  of pure alexia, 794
  of spatial dyslexia, 798
  spatial neglect and, 259, 263–264, 266
  of syntactic processing, 807–814
  in visual plasticity, 131–132
Lettvin, Jerry, 309, 310, 312
Levels of representation, in learning and learning transfer, 159
Levick, Bill, 311
Lewis, Donald, 694
Lexical-semantic systems, 771–773
Lexicon
  infants' early, 843–844
  morphological composition in, 778
Libet, Benjamin, 1192–1193
Libet clock paradigm, 1192–1193
Ligand-gated (NMDA-type) receptors/channels, cross-species comparisons of cortical development, 15
Likelihood function, 527
Limbic system
  neuronal circuitry of frontal lobe, 31
  reward/value signals in, 227–228
Linkage studies
  of developmental dyslexia, 857–858
  genetic linkage sites, 863–864
  nature of, 856
Lipps, Theodore, 974
LISA model of relational reasoning, 1007–1010, 1011, 1013, 1015
Lissencephaly Type I, 14
Little Albert (case), 907
Lømo, Terje, 109–111
Local combination detector (LCD) model, 789–790, 794, 799
Localization error, eye movements and, 514–516
Locke, John, 129
Locked-in state (LIS), 1124–1125
Locus coeruleus (LC), in attachment learning, 891
Locus of attention
  coding in dorsal attention network, 224–226
  spatial attention in visual system, 237–238
Loftus, Elizabeth, 700
Long-term depression (LTD)
  discovery of, 111
  plasticity of synaptic connections, 109–123, 146–147
    overview of cellular mechanisms, 111–113
    properties relevant for memory formation, 113–114
    studies in behaving animals, 114–119
  of visual inputs, 100–101
Long-term memory (LTM)
  consolidation theory and, 692–695
  frontoparietal network and, 229
  molecular and cellular correlates, 697–698
  path integration and, 683–684
  working memory versus, 683
Long-term potentiation (LTP)
  discovery of, 109
  in fear reaction, 912
  hippocampal receptive fields, 119–121
  place fields, 119–121
  plasticity of synaptic connections, 109–123, 146–147
    overview of cellular mechanisms, 111–113
    properties relevant for memory formation, 113–114
    studies in behaving animals, 114–119
  of visual inputs, 100–101
Long-term potentiation in late phase (L-LTP), 112–113, 697–698
Low-frequency stimulation (LFS), 111, 112
LQG, 617, 621–622

M

M cone mosaic
  Bayesian statistics and, 401–404
  described, 395
  functional consequences of, 386–388
  structural organization of, 386, 387
Macaque. *See* Monkeys
MacArthur-Bates Communicative Development Inventories (CDI), 840
Machine consciousness, 1119–1120
Machine learning, 158–159
Macular degeneration (MD), 131–132
Magnetic resonance imaging (MRI), in comparative neuroimaging, 55
Magnetoencephalography (MEG). *See also* EEG/MEG studies
  early language acquisition, 837–839, 842, 847, 850
  empathy and, 975, 982
  neural processing of of emotional information, 925–926, 930
Main complex, 1205
Mammals, cross-species comparisons of neocortical development, 7–23
Manipulation, as cognitive control function, 76
MAP2, 55, 57
Mapping connections, in relational reasoning, 1008
Marginal zone (MZ)
  cross-species comparisons of cortical development, 11, 13
  neuronal circuitry of frontal lobe, 30–32, 33, 43
Marking Paradigm, 193
Marr's theory of object perception, 795
Masking. *See also* Auditory masking
  metacognitive, 1198, 1199
  in saccadic suppression, 512, 1154
Master map of locations (Treisman), 273
Maternal separation/deprivation, 895
Matin, Leonard, 514
Matrix metalloproteinases (MMP), 92
McGurk effect, 499
MCHI (microcephalin), 57–59

Mean population response
dynamics, 425
primate cortex, 422–425
sparseness of, 423–425
spatial spread of, 422–423
variability, 425–427
Measurement density, 527
Medial bank of intraparietal sulcus (MIP), 599–600
Medial entorhinal area (MEA), comparative anatomy of, 667–669, 671
Medial frontal cortex (MFC), in decision making, 1023–1027
Medial geniculate nucleus (MGN), in rewiring vision into auditory pathway, 95–98, 167, 168
Medial nucleus of CM (CEm), in fear reaction, 912–913, 918–919
Medial prefrontal cortex (mPFC)
in conditioned fear reactions, 917, 918
in emotion regulation, 963, 965–966
moral judgment and, 990–991
object concepts and, 1038
social cognition and, 954–956
in surprise reactions, 936–939
Medial superior temporal area (MSTd)
neuron response in variants of heading discrimination, 503–505
optic flow in primates, 502–503
Medial temporal lobe
memory function, 675–687
awareness, 686–687
declarative memory, 686–687
habit learning, 679–680
intact visual perception, 675–678
path integration, 683–684
recollection and familiarity, 680–683
remote memory, 684–686
response to simulation of future events, 755
working memory, 678–679
parahippocampal region, 659–671
boundaries and nomenclature, 660–662
connectivity in, 662–669
damage to, 659–660
defined, 660
entorhinal cortex, 659–660, 661, 662, 667–671
in monkeys, 660–666, 667
parahippocampal cortex, 659, 661, 665–666
perirhinal cortex, 659, 660–665
postrhinal cortex, 663, 666–667, 670
in rodents, 662, 665, 666–671
in perceptual learning, 132
Mediodorsal nucleus, neuronal circuitry of frontal lobe, 35
Medium spiny neuron (MSN), response to severe brain injury, 1130–1131
Memory. *See also* Amnesia; Episodic memory; Priming; Semantic memory; Working memory
autobiographical, 684–686, 726, 731, 753
cognitive control and. *See* Cognitive (executive) control, memory and
cognitive tradition, 691, 699–700
defined, 691
emotional, 700, 725
false memories, 700
imperfections of, 656–657
learning versus, 655
medial temporal lobe function and, 675–687
awareness and memory, 686–687
habit learning, 679–680
path integration, 683–684
recollection and familiarity, 680–683
remote memory, 684–686
visual perception, 675–678
working memory, 678–679
nature of, 655
in perceptual learning, 132–136
physiological tradition, 691
reconsolidation, 656, 691–700
alternative interpretations, 696–698
cognitive implications, 699–700
consolidation theory, 692–693
constraining, 695–696
reviving, 693–695
systems, 698–699
in semantic unification framework, 819
in spatial learning, 114–116, 119–121
synaptic modifications and, 109–111, 113–119, 121–122
Memory impairment. *See also* Amnesia
future-event simulation and, 752–754
Memory processing theory (Lewis), 694
Mental rotation, video games and, 157, 159
Mentalizing, 954, 955–956, 973–974. *See also* Empathy; Social cognition
Messenger RNA (mRNA), comparative, 59–61
Mice. *See* Rodents
Micrographia, striatal damage and, 593–594
Middle longitudinal fasciculus (MdLF), connectivity studies, 261–262
Middle superior temporal (MST)
motion sensitivity of, 166, 318
symmetry and invariance, 318
Middle superior temporal (TST), eye movements and, 513
Middle temporal (MT)
changes with attention, 190
depth perception, 483–496
binocular disparity processing, 484–489
coarse and fine discrimination, 489–493
motion parallax, 484, 493–495
eye movements and, 513
motion sensitivity of, 166, 284–285, 318
symmetry and invariance, 317–318
Midfetal stage
defined, 30
neuronal circuitry of frontal lobe, 30, 31, 32–33, 34
functional organization, 33
laminar shifts in acetylcholinesterase (AChE), 34
neurogenic events, 32–33
structural organization, 32
Midlateral PFC, cascade model and, 1023
Miller, George A., 1221, 1233
Mimicry, 114
Minimally conscious state (MCS), 1123, 1124, 1125–1127, 1130–1134, 1139
Minimum information partition (MIP), in qualia space, 1214
Mirror-invariant representation, 798
Mirror neuron system, 625–638
defined, 625
discovery, 847
human, 633–638
anatomy, 633–635
intention understanding, 635–636
motor cognition in autism, 636–638
plasticity, 635
monkey, 626–633
anatomy, 626
audiovisual, 630–632
functional properties, 627–632
goal-relatedness/chaining, 626–627
and shared brain systems, 847–848
Mismatch negativity (MMN), 840–842, 846
Mismatch problem, and higher order of thought (HOT) approach to consciousness, 114–115
Mnemonic activity, 669–670
reduced interference in memory tasks, 714–715
reduced uncertainty in memory tasks, 715–719
Mnemonic effects of emotion, 730–731
Model-animal approach, comparative approach versus, 52–53
Modes, in qualia space, 1213
Modular hypothesis
of cortical architecture, 1247–1249
in language processing, 1247–1249
Monkeys. *See also* Neural populations in primate cortex
attention research
attentional bottleneck, 282–283
attentional processing, 295, 297
biased competition theory, 198–199
feature-based selection, 241–2492
chimpanzee population decline, 61
comparisons of mammalian cortical development, 7–23, 56–58, 1138–1139
conscious perception in, 153–154
decision making by, 1085–1086, 1089–1090
depth perception in, 489–490
empathy research, 975–982
in evolutionary biology, 49–61
face recognition in, 468, 471, 472
language processing in evolutionary biology, 873–876, 880
memory, analysis of parahippocampal region, 660–666, 667
model-animal versus comparative approach and, 52–53
motor processing, mirror neuron system, 625, 626–633
multisensory integration, 499–508
correlation with behavioral choice, 505–506
cortical neuron response, 502–503
cue reliability, 506–507
heading discrimination, 500–502, 503–505
optic flow, 500–503

Monkeys (continued)
signal detection theory, 503–505
vestibular signals, 500–502
neuronal basis of perceptual illusions, 1143–1144
neuronal circuitry of frontal lobe, 33, 41
plasticity of neurocognition, 165
posterior parietal cortex (PPC), movement anticipation and anticipation, 599–600
visual cortex
formation of eye-specific inputs, 70
grandmother cell studies, 313–315
perceptual learning and, 133, 134–135
task-specific top down influences, 136–137
Monoamine oxidase-A (MAO-A), in imaging genetics studies, 947
Monochromacy
blue cone, 390
rod, 390–392
Monocular deprivation (MD)
partial spatial loss due to, 273–277
and plasticity of ocular dominance, 99–102
Moral judgment, 987–1001
brain damage and, 987–988
dual-process morality, 991–993
mapping moral emotion, 988–991
mental states of moral agents, 993–994
neuroeconomics and, 530, 994–995
Morphemes
derivational, 777
functional, 777–778
ignoring, 779
Morphological processes, 777–785
grammatical categories, 782–785
morphological composition in the lexicon, 778
neural basis of, 778–780
regular/irregular debate in, 780–782
Morphophonology, 780
Morris water maze task, 114–116
Motherese, 846–847
Motion detection, visual plasticity and, 166, 167
Motion parallax
defined, 493
in depth perception, 484, 493–495
Motivation
in avoidance conditioning, 116, 907–909
conditioned, 909
in learning and learning transfer, 161
Motivational control, in decision making, 1023–1027
anatomical definitions, 1024
changing action-outcome associations, 1025
changing task demands, 1024–1025
theories of MFC function, 1024
Motor act, defined, 632
Motor act representations, 627
mirror neural system in monkeys, 626–628
Motor action, defined, 632
Motor cortex (M1)
athletic domain, 155, 156–157
cortical mapping studies, 142–146
forward state estimation for online control, 602–603
organization, 553, 554–555
specificity of learning, 154, 155–156
Motor effector manipulations, in phonological processing, 771
Motor primitives, 541–549
compositionality, 543, 547–548
modularity in frog spinal motor system, 546–547
movement planning to execution, 541–543, 547
nature of, 543
synergies, 543–546, 548–549, 619–621
Motor processing
in autism, 636–638
basal ganglia function and, 568–573, 575–577, 579
computational neuroanatomy, 581–596
cerebellum, 591–592
limitations in applying theory, 595
parietal cortex damage and state estimation, 594
problem in reaching, 589–590
problem of motor control, 587–588
rewarding nature of sensory states, 592–593
striatal damage in assessing movement costs and rewards, 593–594
consciousness and
conscious veto of, 1193–1194
spontaneous motor initiation, 1191–1193
mirror neuron system, 625–638
defined, 625
human, 625, 633–638
intention understanding, 632–633, 635–636
monkey, 625, 626–633
object concepts, 1035
optimality in, 615–617
sensory processing compared with, 613–622
algorithms for learning and online computation, 621–622
Bayesian inference, 614–622
inference versus control, 613–614
intermediate representations, 619–621
optimal control, 614–619
Motor speech areas, in phonological processing, 770–771
Motor-speech systems, 771–773
Motor theories of perception, 773
Movement release and inhibition
basal ganglia function, 568–573, 579
conscious, 1194–1195
Mozart effect, 154–155
MUC (memory, unification, control) framework, 819
Multidimensional scaling (MDS), in auditory object analysis, 368–370
Multiple-self models, in valuation process, 1088–1089
Multiscale spatial frequency filing theory, 436
Multisensory integration, 499–508. *See also* Sensory processing
correlation with behavioral choice, 505–506
cortical neuron response, 502–503
cue reliability, 506–507
heading discrimination, 500–502, 503–505
optic flow, 500–503
in phonological processing, 770–771
signal detection theory, 503–505
vestibular signals, 500–502
Muratoff's fascicle, 36
Muscle coordinates, 542
Muscle synergies, motor behavior and, 543–546, 548–549, 619–621
Musical domain
effect of music lessons on IQ, 156
Mozart effect, 154–155
Myelination, of white matter, 76
Myelinization, neuronal circuitry of frontal lobe, 42
Myelogenesis, neuronal circuitry of frontal lobe, 39, 41
Myohyloid muscle (MH), in autism, 636–638

N

N170 potential, in face recognition, 472–473, 474–477, 477
N200 potential, neural signatures of word learning, 843
N290 potential, in face recognition, 472–473
N400 potential, in semantic processing, 169, 170, 820–827, 845
Narrative self, 954
National Institutes of Health (NIH), 61
Native language neural commitment (NLNC) hypothesis, 840, 850
Near-infrared spectroscopy (NIRS), early language acquisition, 837–838, 839, 846–847
Neglect dyslexia, 795–796
Neocortical function
cross-species comparisons of mammals, 7–23
cognitive processing, 7–8, 20, 22–23
cortical architecture, 7
cortical size determinants, 15–17
differentiation level, 52–53
neuronal cell migration, 10–14
onset of neurogenesis, 8
origins of cortical neurons, 8–9
protomap hypothesis, 18–19
radial unit hypothesis, 14–15
span of neurogenesis, 8
stages of cortical development in humans, 20–22
synaptic connections, 7, 19–20
transient embryonic zones, 9–10
prefrontal. *See* Prefrontal cortex (PFC)
Neonatal stage
defined, 30
neuronal circuitry of frontal lobe, 30, 37–39, 44
functional organization, 37–39
neurogenetic events, 37
structural organization, 37

Neural networks
computational model for texture filling-in, 442–447
global architecture, 442–443
single processing units, 442
dynamic changes in activity, 142–146
frontoparietal attention network, 219–227
network-brain links, 443–444
perisylvian, 259–266
synaptic plasticity in, 121–123
Neural populations in primate cortex
decoding, mechanisms, 427–429
encoding, 419–432
mean response, 422–425
response variability, 425–427
theoretical framework, 420–421
in vivo measurement, 421–422
Neural priming, 715, 717
Neural signatures
in sentence processing, 844–845
in typically developing children, 839–841
of word learning, 842–843
Neural tuning, 456–463
central and anterior (CIT/AIT), 461–465
posterior (PIT), 458–461, 463–464
Neuroeconomics
defined, 530, 1075
emotion in, decision making and, 1098–1100
of goal-directed choice, 1075–1083
computational basis, 1079–1081
multiple behavior controllers, 1078–1079
neurobiological basis, 1079–1083, 1095–1100
simple binary stimulus choice, 1076–1077, 1079–1083
moral judgment and, 994–995
valuation in, 1085–1091
amygdala and, 1097
basic structure, 1087–1089
choice and, 1089–1091
initial investigations, 1098–1100
striatum and, 1097
two-stage model, 1085–1087
Neuroepithelium, cross-species comparisons of cortical development, 9–10
Neuroimaging. *See also* Imaging genetics; *specific types of imaging*
comparative, 55–58
emotion regulation and, 961–970, 1099
commonalities across strategies, 968–969
differences across strategies, 969
future directions, 969–970
unique executive function system, 969
functional, 58
of future-event simulation, 754–757
individual differences in, 741–742
of morphological processing, 779, 780–785
random-effects analysis, 742
structural, 55–58
Neuronal cell migration, cross-species comparisons of cortical development, 10–14
Neuronal correlates of consciousness (NCC), 1140–1141, 1143–1145, 1172–1176
Neuronal mechanisms, in perceptual learning, 133–136
discrimination learning, 133–134
enhanced responsiveness, 134–135
shifted cortical representation, 134–135
Neurons. *See* Neural populations in primate cortex; *specific types of neurons*
Neuropeptides
neuronal circuitry of frontal lobe, 35
neuropeptide Y (NPY) in imaging genetics studies, 948
Neuroplasticity, 141–150, 165–176. *See also* Plasticity
as cause of disease, 147
complementary mechanisms in control of, 146–147
cortical mapping studies, 142–146
cross-species comparisons of cortical development, 19
interventions, 148–149, 173–176
of mirror neural system, 635
as opportunity for intervention, 148–149, 173–176
plasticity as normal state, 142
profiles
attention, 170–173
audition, 167–169
language, 169–170
vision, 166–167
Neurotransmitters, cross-species comparisons of cortical development, 18
NLM-e, 844
NMDA (*N*-methyl-D-aspartate) receptors
muscle synergies and, 546–547
synaptic plasticity and, 111–116, 117, 119–121
No consciousness theory of consciousness, 1114
Noise, additive versus multiplicative, 425
Noise exclusion, attention and, 171, 173
Nondeclarative memory, forms of, 675
Nonspatial working memory, development of, 75
Nonspeech signals, 846–847
Noradrenaline, 1140
Norepinephrine, 1140
Norepinephrine (NE), in mediating learned behavior, 890–891, 894
Notch functions, cross-species comparisons of cortical development, 9
Novel orientation, in learning and learning transfer, 159
NPY, 35
Nucleus accumbens (NAcc)
in conditioned fear reactions, 917–918
in decision making, 1094, 1097
Numb (protein), cross-species comparisons of cortical development, 9
Nyquist frequency, 383

O

Object analysis
auditory. *See* Auditory processing
concept of, 367–368
Object-based selection, 191–192
attentional-biasing signals, 208
loss in Balint's syndrome, 270–271
neuroimaging studies, selecting among multiple competing objects, 210–214
physiological mechanisms, 229–230
in spatiotemporal analysis of visual attention, 240–246
units of selection, 208, 209
Object classification, 159
Object concepts
defined, 1031
hierarchical organization of, 1031–1032
nature of, 1031–1032
neural foundations, 1032–1042
acting, 1032–1035
anterior regions of temporal lobe, 1040–1041
domain-specific neural circuitry, 1035–1041
feeling, 1032–1035
perceiving, 1032–1035
visual processing, 1032
Object recognition
auditory, 367–380
concept of auditory object, 367–368
fMRI studies, 371–379
manipulation of natural stimuli, 368–370
stimuli based on sequence of objects, 370–371
synthetic stimuli, 370
in grasping studies, 644
visual, 159, 455–465
in discriminative learning, 133–134
network processing of boundary fragments, 463–464
object concepts and, 1031–1042
retinal representation, 455
three-dimensional surface fragments, 460–463
two-dimensional object boundary fragments, 455–460
word perception as object perception, 789–792
Object substitution masking, 200
Objective approaches, to measuring visual awareness, 1151–1152
Observer framework, 601–603
accounting for neurobiological explanations, 1205–1209
feedback in visual attention, 1165–1175
ideal Bayesian observer, 420–421, 530
Kalman filter, 602–603, 608–609
nature of observer, 602
sensory biases in selective attention, 219–223
Obsessive-compulsive disorder (OCD), 561
basal ganglia function and, 568, 571, 572, 577
Occipitotemporal sulcas, 662
Ocular dominance
ocular dominance columns (ODCs)
eye-specific inputs to dLGN, 68–70, 94
formation of, 67–68
plasticity of, 91, 99–102

Ocular dominance (continued)
anatomy of, 99
changes in extracellular matrix, 101
critical periods, 100
gene screens for novel factors, 101–102
Hebbian mechanisms, 100–101
homeostatic mechanisms, 100–101
lid suture, response to, 99–100
long-term depression (LTD) of inputs, 100–101
long-term potentiation (LTP) of inputs, 100–101
monocular deprivation (MD), 99–102
structural, 101
Odor learning
in attachment, 889–890
in aversion, 892–893
Olfaction, 321–336
encoding
odor, 326–327, 330–332
from percept to molecule, 330
spatial, 327–329
temporal, 329–330
infant odor learning and attachment, 889–890
odor space
constructing perception-based, 330–332
model from physical to perceptual space, 333
physiochemical molecular descriptor space, 332–333
to predict neural activity in olfactory system, 333–335
olfactory system
described, 322–324
olfactory bulb, 325–326, 328
olfactory cortex, 326
olfactory epithelium, 324–325, 327–328
predicting neural activity in, 333–335
Olfactory cortex
described, 326
infant odor learning and attachment, 889–890
Open loop control, 615–617
Operant conditioning, 310
Optic chiasm formation, 92
Optic flow, 318
perception of heading from, 500–502
primate cortical neuron response to, 502–503
Optic tectum, targeting and retinotopic wiring, 92
Optimal control theory, 549
duality with Bayesian inference, 614–622
algorithms for learning and online computation, 621–622
general duality, 617–619
intermediate representations, 619–621
isometric tasks, 614–615
optimality in sensory and motor processing, 615–617
in motor learning and control, 588
optimal estimation versus, 618–619
Optimal decoder, 420
Optimal encoder, 420
Optimal estimation, 525–533
Bayesian formulation, 527–529, 530–531
in the brain, 530–533
computational basis for motor synergies, 548–549
defined, 525
formulations of, 525–529
optimal control versus, 618–619
physiological implementation, 531–533
regression formulation, 525–527
Optimal lag time (OLT), 604–609
Orbitofrontal cortex (OFC)
in conditioned fear reactions, 917, 918
in emotion regulation, 965, 966–968, 969
moral judgment and, 988
in processing of emotional information, 932
in simple binary stimulus-choice paradigm, 1081–1083
Order discrimination, in relative-learning tasks, 359
Orexin, 1140
Orientation discrimination, in discriminative learning, 133–134
Orientation scotomas, 411–413
Orientational selectivity, 309, 311–312
Origin of Species (Darwin), 50–52
Other-race effect, in face recognition, 468
Overgeneralization, 1048, 1051–1052, 1059

P

P600 potential, in semantic unification, 824–826, 845
Pain
awareness of, in infancy, 1117–1118
comparisons of species, 1138
Pain syndromes, 147
Pajek software, 1224, 1226–1227
Parahippocampal cortex
comparative anatomy of, 659, 665–666
damage to, 659–660
Paralexias, 779
Parallel distributed processing (PDP), 1047–1053
in conceptual development, 1048–1053
in conditioned fear reactions, 913–914
fundamental tenets, 1047–1048
Paralysis, transient form of, 1141
Parent training, 176
Parietal cortex
audition and, 167–168
frontoparietal attention network, 219–230
anticipatory signals, 219–223
coding locus of attention, 224–226
establishing goals, 227–229
eye movements, 226–227
feature-based selection, 226–227
functional connectivity, 219–223
physiological mechanisms for selection of visual objects, 229–230
prefrontal-cingular circuits, 228–229
reward/value signals and limbic system, 227–228
task sets, 228–229
top-down biases, 223, 251–252
topographic organization of maps, 223–226
working memory, 228–229
frontoparietal mirror neuron system in motor processing, 625–638
defined, 625
human, 625, 633–638
intention understanding, 632–633, 635–636
monkey, 625, 626–633
impact of damage on state estimation, 594
stroke in, 144
Parkinson's disease, 560–561
basal ganglia functions and, 565, 567, 568, 570–574, 577, 594
effects of striatal damage, 593–594
Parsing, 812, 813
Part-in-spacing-altered-whole effect, in face recognition, 468
Part-whole effect, in face recognition, 468
Path integration, medial temporal lobe function, 683–684
Pathological laughter and crying (PLC), 1184
Pavlovian conditioning, 310
actions in, 906–909, 914–918
avoidance conditioning, 116, 907–909
components of, 906–909
fear response, 906–919
goal-directed choice and, 1078–1079, 1083
learning tasks, 116, 118, 319
Pavlovian-to-instrumental transfer (PIT), 909, 915–917
Pax6 genes, 876–877, 878, 879
People Pieces Analogies, 1011
Perception. *See also* Depth perception; Surface perception
basal ganglia function and, 575–577
emotion and, 925–932
automaticity, 929–931
behavioral effects on perception, 926–928
cultural factors, 931–932
current behavior goals, 932
neural circuits underlying attention, 928–929
neural processing in perception, 925–926
personality factors, 931–932
knowing versus, 1035
learning versus, 675
motor theories of, mirror neurons and, 773
neuronal basis of perceptual illusions, 1143–1145
object concepts and, 1031–1042
Perception-based odor space
constructing, 330–332
in predicting neural activity in olfactory system, 333–335
Perceptual asymmetry, in early visual processing of printed words, 791–792
Perceptual face-space, 468
Perceptual filling-in, 435–449
active interpolation theory
described, 435
early visual system, 437–439
evidence for, 439–442

computational model, 442–449
defined, 436–437
future research directions, 449
insights, 447–449
limitations of research, 449
processes in surface perception, 435–441
simulation studies, 443–447
Perceptual learning, 132–136
auditory, 132–133, 353–363
characteristics, 358–362
neural processes, 353–358
neural underpinnings, 362–363
cortical recruitment in, 132–133
neural underpinnings of, 362–363
neuronal mechanisms, 133–136
psychophysics of, 132
Perceptual masking, 348–349
Perceptual memory, executive memory versus, 708–709
Perceptual narrowing, in face recognition, 471–472
Perceptual organization principles, 212–213, 317
Perceptual theory of speech production, 773
Perceptual valence, 331–332
Perforant path (PP), 109
Performance monitoring, as cognitive control function, 76
Perfusion-weighted imaging (PWI), connectivity studies, 264
Periaqueductal gray (PAG), 878
Perineuronal nets (PNN), in ocular dominance plasticity, 101
Perinuclear cage, cross-species comparisons of cortical development, 14
Perirhinal cortex, comparative anatomy of, 660–665
Perisylvian neural network, 259–266
role in human right hemisphere, 263
spatial neglect as disconnection syndrome, 263–264
white matter connectivity, 260–262
Personality factors, in processing of emotional information, 931–932
Phenomenal consciousness. *See also* Higher order of thought (HOT) approach to consciousness
during childhood, 1117–1118
described, 1113–1114
Phenomenology, in qualia space, 1214–1215
Phenotype, discovery and changes, 59–61
Phenotypes, 876
Phobias, 927, 928
Phonemes
mirror system for, 625
in second-language learning, 842
Phonetic learning
interaction with word learning, 844
neural signatures in typically developing children, 839–841
Phonological processing, 767–774
auditory-motor integration networks in, 771
defined, 767
left posterior planum temporale (area Spt) in, 770–771
mirror neurons in, 773
morphological processing and, 779
motor theories of perception in, 773
posterior language cortex in left hemisphere, 770
speech recognition, 769–773
spoken word recognition, 768–769
superior temporal sulcus (STS) in, 769
task dependence, 767–768
visual representations and, 799
Photodiode thought experiment, 1202–1203, 1211–1212
Phylogenetic scale, 50
Physiochemical molecular descriptor space, 332–333
Piaget, Jean, 78
Pictorial cues, 483
Place cell networks, 121–123
Place fields, in long-term potentiation (LTP), 119–121
Plaid paradigm (Adelson and Movshon), 317–318
Planum temporale
in human brain specialization, 53
in phonological processing (area Spt), 770–771
connection to motor speech areas, 770–771
motor effector manipulations, 771
sensorimotor response properties, 770
speech related visual stimuli, 771
Plasticity. *See also* Neuroplasticity
amygdala, lack of, 892
defined, 89
dynamic activity across neural networks, 142–146
as normal state, 142
synaptic, 109–123
attractor dynamics in neural networks, 121–123
long-term depression (LTD) of cellular mechanisms, 109, 111–119, 122–123, 146–147
long-term potentiation (LTP) of cellular mechanisms, 109, 111–121, 122–123, 146–147
modifications as means for memory, 109–111
training-related, 153–161
complex learning environments, 156–158
determinants of learning and learning transfer, 158–161
impact of practice, 154–156
specificity of learning, 153–154
visual cortex, 129–137
lesions and, 131–132
neuronal activity in formation of eye-specific connections, 67–71
ocular dominance plasticity, 91, 99–102
perceptual learning and, 132–136
postnatal development, 129–131
profiles, 166–167
top-down influences, 134, 136–137
Pleasantness, perceptual, olfactory, 331–333, 336
Point of view, 1203, 1205
Populations. *See* Neural populations in primate cortex
Positron emission tomography (PET)
basal ganglia function and, 576
brain function and, 1067–1068
comparative, 57
in emotion regulation, 964, 970
forebrain dysfunction following severe brain injury, 1130
motor cortex activity, 142
neural processing of emotional information, 925–926
object concepts and, 1033, 1040
posterior parietal cortex (PPC) and, 601
in semantic memory processing, 1062
in semantic unification network, 831
simulation of future events, 754–756
Posner's cuing paradigm, 236, 242, 282
Postconceptual week (PCW), in stages of corticogenesis process, 29, 30–37
Postdiction, eye movements and, 518
Posterior cingulate cortex (PCC)
moral judgment and, 988, 993–994
reward/value signals in limbic system, 228
Posterior IT (PIT) neural tuning, 458–461, 463–464
Posterior language cortex, in speech production, 770
Posterior lateral PFC, cascade model and, 1022–1023
Posterior parietal cortex (PPC), 167
state estimation, 599–609
dynamic, 606–609
movement intention and anticipation, 599–601
neural correlates, 604–607
for online control, 601–603
sensorimotor control, 603–607
virus tracing studies, 558
Postmitotic cells, cross-species comparisons of cortical development, 18–19
Postphonemic processing, 769
Postrhinal cortex, comparative anatomy of, 663, 666–667, 670
Posttraumatic stress disorder (PTSD), 147, 928
Power spectrum model of masking, 345
PR-LTM (postreactivation long-term memory), 693–698, 700
PR-STM (postreactivation short-term memory), 693–698, 700
Preattentive indices (FINSTs), 192
Precision grip, mirror neural system and, 626–638
Precognition period, neuronal circuitry of frontal lobe, 29
Predictive remapping, eye movements and, 517–518
Prefrontal cortex (PFC)
in cognitive control development, 76–77, 82, 705–709, 714–719, 1019–1027
comparative neuroimaging, 57–58
development stages, 29–44
in emotion regulation, 963–969
in fluid reasoning development, 79–80
modularity of language and, 1238

Prefrontal cortex (PFC) (continued)
in relational reasoning, 1008–1015
social cognition and, 956–957
structural brain development, 73–74
virus tracing studies, 557–558
Premotor areas, 555–556
Premotor cortex
cascade model and, 1022
goal-directed actions and, 1037–1039
Premotor theory of attention, 191
Preparatory responses, 914
Preplate (PP), cross-species comparisons of cortical development, 11
Presubplate (PSP), neuronal circuitry of frontal lobe, 30, 31, 32
Presupplementary motor area (PreSMA), virus tracing studies, 558
Primary motor cortex. *See* Motor cortex (M1)
Primates. *See* Monkeys; Neural populations in primate cortex
Prime order of thought (POT), explanatory gap and, 1114
PRIMER, 844
Priming
attention and, 196, 197, 199
negative, 196, 197
partial spatial loss, 274–276
conceptual, 715–719
for early language acquisition, 840
neural, 715, 717
Primordial plexiform layer, neuronal circuitry of frontal lobe, 29–31
Principle components analysis (PCA), 330–332
Principle of delayed estimating, 533
Principle of perceptual organization, 212–213, 317
Prisoner's dilemma game, 995
Procedural/declarative hypothesis, in morphological processing, 780–782
Processing, competence versus, 1241–1242
Programmed cell death (PCD), cross-species comparisons of cortical development, 16
Proliferative zones. *See also* Subventricular zone (SVZ); Ventricular zone (VZ)
cross-species comparisons of cortical development, 14
neuronal circuitry of frontal lobe, 34–35, 36, 42
Propositions, in syntactic processing, 805–806
Proprioception, 601, 604, 606
Proteases, in ocular dominance plasticity, 101
Protein kinase C (PKC), 116
Protein kinases, 111
Protein synthesis, in ocular dominance plasticity, 100–101
Protein-synthesis inhibitor, 698
Protomap hypothesis
cross-species comparisons of cortical development, 18–19
neuronal circuitry of frontal lobe, 30–31
Psychological refractory period, for attention, 194, 195
Psychopathology
empathy and, 974, 979
future-event simulation and, 753–754
moral judgment and, 988
Psychopathy
empathy and, 974, 979
moral judgment and, 988, 993
Psychopathy Checklist, 993
Punishment, 967–968
Purdue Pegboard task, 149
Pure alexia, 794
Pure autonomic failure (PAF), empathy and, 977
Purkinje cells, 1140
Put, putamen, neuronal circuitry of frontal lobe, 31
Pylyshyn, Zenon, 1241–1245

Q

Qualia space (Q-space), 1212–1216
defined, 1212
giving shape to experience, 1212–1213
translating phenomenology into geometry, 1214–1215
viewing as shapes, 1213–1214
Quantum mechanics
consciousness and, 1141
SQUID (superconducting quantum interference device), 839

R

Radial domain, neocortical development, cross-species comparisons, 7, 9, 13, 14–15, 22, 23
Radial glia (RG)
cross-species comparisons of cortical development, 7, 9, 13, 14–15, 22, 23
neuronal circuitry of frontal lobe, 33
Radial wait hypothesis, cross-species comparisons of cortical development, 14–15
Ramón y Cajal, Santiago, 109, 129
Random-effects analysis, 742, 744, 745
Rape, 928
Rapid auditory processing, 168
Raven's Progressive Matrices (RPM), 78–81, 1008–1009, 1015
Reach adaptation, in motor learning and control, 589–590
Reaction times, video games and, 157
Reactions
fear, neural basis for, 909–914
nature of, 906
in Pavlovian conditioning, 906–914
Reading ability, 175–176
Reading impairment, 166
Reading intervention, 174–176
Reading process, 789–800
dorsal visual pathway, 794–799
interfacing with verbal system, 798–799
orientation of attention, 795–796
parts of words, 797–798
pathology, 795–796, 797, 798
serial decoding, 797–798
single word selection, 796–797
pathology in, 792–798
signed language, 169–170
ventral visual pathway, 789–794
early visual processing of printed words, 791–794
interfacing with verbal system, 798–799
pathology, 792, 794
word perception as object perception, 789–792
Reafference, posterior parietal cortex (PPC) and, 600–601
Reagan, Ronald, 1260
Reappraisal, in emotion regulation, 964–966, 969
Receiver operating characteristics (ROC), 680–683
Receptive fields (RF)
population properties, 422–423
in primary visual cortex, 409–411
Recipient, in relational reasoning, 1008
Recollection, medial temporal lobe function, 680–683
Reconsolidation, 656, 691–700
alternative interpretations, 696–698
cognitive implications, 699–700
consolidation theory, 692–693
constraining, 695–696
constraints on, 695–696
reviving, 693–695
systems, 698–699
Red/green color vision deficiency, 389–390
Reed, Randy, 324
Reflexes, nature of, 905
Regional selectivity, in Visual Word Form Area (VWFA), 792–794
Regionalization, in formation of visual pathway, 91–92
Relational complexity, 1008–1010, 1013–1014
Relational integration, 1006–1007, 1012–1013
Relational reasoning, 1005–1015
component processes, 1006–1008
interference control, 1006–1007, 1013–1014
LISA model, 1007–1008, 1011, 1013, 1015
relational integration, 1006–1007, 1012–1013
prefrontal cortex (PFC) in, 1008–1015
aging process, 1010–1011
neuroimaging evidence, 1011–1015
neuropsychological evidence, 1008–1011
relational complexity, 1008–1010, 1013–1014
Relative entropy, 1204
Repetition suppression (RS) phenomenon, 647–648, 715–719
Repetitive transcranial magnetic stimulation (rTMS)
morphological processing, 782–785
motor cortex activity, 142–146, 148–149
Representation. *See also* Conceptual representation

levels of, in learning and learning transfer, 159
in phonological processing, 767–768
in simple binary stimulus-choice paradigm, 1081
syntactic processing, 805–807
visual, consciousness and, 1112, 1113
Reproducible localization, in Visual Word Form Area (VWFA), 792
Repulsive bias, 531
Repulsive guidance molecule (RGM), 94
Response inhibition. *See also* Long-term depression (LTD)
as cognitive control function, 76–77, 78
conscious, 1194–1195
movement release and, 568–573, 579
nature of, 76–77, 78
Response selection
as cognitive control function, 76–77, 78
nature of, 76–77, 78
Response variability of population, 425–427
additive versus multiplicative noise, 425
spatial correlation, 425–426
temporal correlation, 425–427
Reticular activating system, 1140
Retinal development, formation of eye-specific inputs, 67–71
Retinal ganglion cell (RGC)
projections to dorsal lateral geniculate (dLGN), 67–71
receptive fields in visual cortex, 410–411
in rewiring vision into auditory pathway, 97
statistical connectivity theory, 411–415
targeting and retinotopic wiring, 92–94
Retinal Motion condition, 495
Retinal processing. *See also* Color vision
lesion studies, 131–132
neurodegenerative diseases, 131–132
retinal waves in, 68, 70–71
visual object recognition, 455
Retinitis pigmentosa, 389
Retinotopic processing, in early visual processing of printed words, 791–794
Retinotopic projections/maps, 91–102
activity-dependent refinement of, 98–99
of dorsal attention network, 223–226
eye movements and, 520–521
formation of visual pathway in early development, 91–95
eye-specific domains, 94
new maps, 94–95
other feature maps, 94–95
regionalization, 91–92
retinotopic wiring, 92–94
targeting, 92–94
ocular dominance plasticity, 99–102
rewiring vision into auditory pathway, 95–98
Retinotopy, surface perception and, 437–438, 442
Retrieval. *See also* Memory
emotional modulation during, 731–733
episodic, individual differences and, 743–747
neural basis of, 743
Retrograde alteration, of synaptic plasticity, 114, 116
Retrograde amnesia, 684–685
Retronasal olfaction, 324
Retroviral gene transfer method, cross-species comparisons of cortical development, 9, 14–15
Reverse hierarchy theory (Hochstein and Ahissar), 194
Rewards
in decision making, 1094–1095
immediate and remote relevance of, 1026
in learning process, 161, 588, 592–593
in motor learning and control, 588, 592–593
reward/salience signals in basal ganglia, 561, 571–572, 575
reward/value signals in limbic system, 227–228
in valuation process, 1087
Rhesus monkey. *See* Monkeys
Right visual field (RVF)
in early visual processing of printed words, 791–793
in semantic processing, 820
Rod monochromacy, 390–392
Rodents
auditory cortex, rewiring vision into, 95–98
basal ganglia function and, 575–577
cross-species comparisons of mammalian cortical development, 7–23, 1138–1139
infant, attachment learning, 890, 895
learning rewarding nature of sensory states, 593
memory, analysis of parahippocampal region, 662, 665, 666–671
model-animal versus comparative approach and, 52–53
olfactory cortex, 316, 329–330
synaptic plasticity, 114–119, 122–123
visual cortex
formation of eye-specific inputs, 68, 70
plasticity, 100–101
visual pathway formation, 91–95
Rosch, Eleanor, 1031
Rostrolateral PFC (RLPFC), development of, 74, 79, 80–81, 707–709
Rule/task-set representation, as cognitive control function, 76
Rumelhart model, 1048–1051, 1055, 1059
Russell, Bertrand, 1116
Ryk receptor expression, 94

S

S cone mosaic
Bayesian statistics and, 401–404
described, 395
functional consequences of, 384–386
structural organization of, 384
Saccades. *See* Eye movements, saccades
Salience assignment, basal ganglia function and, 573–575
Sampling methods, 622
Scene parameters, in color vision, 396
Schaffer collateral pathway, 110, 111
Scharff, Constance, 879
Schizophrenia, 147
basal ganglia function and, 571
visual awareness and, 1156
Second-language learning, 169, 842, 848–849
Segment III of von Monakow, 36
Selection negativities, 238–240
Selection positivities, 238–240
Selective attention. *See also* Attentional processing, in selective attention
auditory, 171–174
in auditory informational masking, 345–348
as cognitive control function, 76
degree of competition, 186
event-related potential (ERP), 186
feature-based selection, 197–198
in analysis of visual attention, 238–240
eye movements in, 226–227
feature integration theory, 197–198, 208–210, 226–227, 276–277
nature of, 76, 185
neuroimaging studies
selection among multiple competing objects, 210–214
units of selection, 206–210
through neuronal synchronization, 289–299
attentional processing, 293–298
attentional selection in neuronal communication structure, 290–292
interneuron networks and attentional modulation, 292–293
selective synchronization, 292
object selection
attributes, 191–192
among multiple competing objects, 210–214
nature of object-based attention, 208–209
physiological mechanisms, 229–230
spatiotemporal analysis, 240–246
plasticity and, 170–171
selectivity of visual neurons, 284–285
spatial selection, 191–192, 206–208
focus of attention, 235–237
locus of attention in visual system, 237–238
spatiotemporal analysis, 235–238
visual search, 238
visual. *See* Visual attention
Selective Tuning Model, 236
Self-awareness
self-knowing, 752
in social cognition, 954–955
Self-consciousness, 1141–1142
Self-projection, 757–758
Self-regulation, in social cognition, 956–957
Semantic aphasia, 1063–1064

Semantic cognition, 1047–1064
neural basis of, 1061–1064
object concepts and, 1040–1041
parallel distributed processing (PDP) and, 1047–1053
modeling semantic dementia, 1054–1057, 1059
overgeneralization, 1048, 1051–1052, 1059
sensitivity to coherent correlation, 1052–1053
Rumelhart model, 1048–1051, 1055, 1059
in semantic aphasia, 1063–1064
in semantic dementia, 1008, 1047, 1053–1061
Semantic dementia, 1008, 1047, 1053–1061
characteristics of, 1054
development versus disintegration and, 1057–1059
knowledge of both objects and words, 1060–1061
modeling, 1054–1057
nonsemantic deficits in, 1059–1060
role of anterior temporal lobes (ATL) in semantic memory, 1062–1063
Semantic facilitation index (SFI), 1010
Semantic memory
loss of, 79
object concepts and, 1033, 1035
role of anterior temporal lobes (ATL) in, 1062–1063
Semantic processing, 169, 170
in action semantics, 648–650
central feature, 875
challenge of, 879–880
interactions between control and memory, 713–714
in relational reasoning, 1007–1008
semantic unification, 819–833, 1062–1063
beyond sentence level, 828–832
functional characteristics, 820–827
integration versus unification, 832
MUC (memory, unification, control) framework, 819
multimodal nature, 828
Sensitive periods. *See* Critical (sensitive) periods
Sensitivity to spacing changes, 468
Sensorimotor control, 603–607, 770, 1021
Sensory processing. *See also* Auditory processing; Multisensory integration; Olfaction; Visual processing
attention and strength of response, 283–284
emotional impact on, 727–728
motor processing compared with, 613–622
algorithms for learning and online computation, 621–622
Bayesian inference, 614–622
inference versus control, 613–614
intermediate representations, 619–621
optimal control, 614–619
neuronal circuitry of frontal lobe, 42
object concepts, 1034–1035
optimality in, 615–617
Sensory substitution
audition for vision, 167–169, 321
in blindness, 321
Sentence processing
in early language acquisition, 844–845
interpretation, 812, 813
semantic unification beyond sentence level, 828–832
Sentience, 1138–1139
Sequencing of candidate genes, nature of, 856
Sequential finger movement task, 149
Serial behavior, hierarchical processing, 641–642
Serial decoding
in conditioned fear reactions, 911–913, 914
in reading process, 797
Serial reaction time (SRT), basal ganglia function and, 576, 577
Serotonin, 1140
Serotonin transporter (5-HTTI), in imaging genetics studies, 947
Short-term memory (STM). *See* Working memory
Sign rule of connectivity, 412
Signal detection theory, 134, 503–505
recognition memory, 680–683
Signal-to-noise ratio, 285–286, 310, 345, 363
Signaling centers, in formation of visual pathway, 91–92
Signed language, 169–170
Silent meaning, in semantic unification framework, 824
Simple binary stimulus choice, 1076–1077, 1079–1083
Simulation studies
future event simulation, 752–754, 757–759
in integrated information theory (IIT), 1207–1209
of perceptual filling-in, 443–447
Simulation theory, 1185
Single word selection, in reading process, 796–797
Size of cerebral cortex, cross-species comparisons of cortical development, 15–17
Skin conductance response (SCR), autonomic response to nonconscious emotion, 1186–1187
Skinner, B. F., 837
Skinner box, 1075, 1077
Sleep
consciousness during, 1139, 1141, 1209
in memory consolidation, 730–731
Slow activity transients (SAT), 35
Smell. *See* Olfaction
SMI-32 (NPNF), 55, 56, 57
Smid, John Maynard, 876
Smith, G. Elliot, 315–316
SOAR, 159
Social cognition, 953–959, 1258. *See also* Empathy
components of, 954–958
detection of threat, 957–958
mentalizing, 954, 955–956
self-awareness, 954–955
self-regulation, 956–957
emotion and, 1184–1185. *See also* Emotion
nature of, 953–954
"special" nature of social cognition, 958
Socioeconomic factors
intervention style and, 176
IQ and, 80, 171
Sociopathy, moral judgment and, 988
Somatic marker hypothesis, 1098
Somatic markers, 977
Somatosensory cortex
neuronal circuitry of frontal lobe, 35
in perceptual learning, 132–133
Somatostatin, 35
Source amnesia, 1118
Space-time tuning function (STTF), 604–607
Spatial attention
comparison of effects across visual system, 208, 1156–1157
deficits in
complete loss, 270–273
loss of perceptual space, 269–273
partial loss, 273–277
types of deficits, 269–270
implicit spatial maps, 272–273
locus of spatial selection, 237–238
orientation in reading process, 795–796
profile of spatial focus of attention, 235–237
role of spatial selection in visual search, 238
spatial neglect versus, 259, 263–264
spatial selection and
behavioral coherence, 191–192
spatiotemporal analysis, 235–238
units of selection, 206–208
in spatiotemporal analysis of visual attention, 235–238
Spatial correlations, in population response, 425–426
Spatial dyslexia, 798
Spatial encoding, in odor discrimination, 327–329
Spatial neglect, 259, 263–264, 265
Spatial orientation
cortical areas in, 259
perisylvian neural network, 259–266
role in human right hemisphere, 263
spatial neglect as disconnection syndrome, 263–264
white matter connectivity, 260–262
Spatial tasks, auditory learning in, 359–362
Spatiotopic maps, 520–521
Species-typical behaviors, 906
Specific language impairment (SLI), 168, 171, 172, 174, 859–860, 863, 864
Specificity
of consciousness, 1209–1212
of input in memory formation, 113, 114
of learning, 153–156
Speech arrest, 260
Speech production
mechanisms underlying, 878
motor-speech systems in, 771–773
neuronal circuitry of frontal lobe, 41

perceptual theory of, 773
posterior language cortex in, 770
Speech recognition
informational masking and, 348–351
motor theory of speech perception, 883
phonological processing
bilateral/asymmetric nature of, 769–770
bilateral organization, 768–769
Speech segmentation, 169
Speech sound disorder (SSD), 860, 863, 864
Spindle cells, in human brain specialization, 54–55
Spinogenesis
first phase, 37
neuronal circuitry of frontal lobe, 37, 41, 42
Spinoza, Baruch, 1183
Split-brain procedures
modular hypothesis and, 1247–1249
word recognition and, 768
Spors, Hartwig, 329
Sprague effect, 1070–1071
Spreading activation
anatomical basis for, 437–438
defined, 436
early visual system, 437–438
SQUID (superconducting quantum interference device), 839
Stanford Binet IQ, 154–155
Starburst amacrine cells, 69
State-dependent learning, 137
State estimation
in motor learning and control, 588, 594
posterior parietal cortex (PPC), 594, 599–609
dynamic, 606–609
movement intention and anticipation, 599–601
neural correlates, 604–607
for online control, 601–603
sensorimotor control, 603–607
Steady-state visual-evoked potential (SSVEP), 236, 239, 240
Sternberg, Robert J., 78
Stimulus-independent thoughts (SITs), 1070
Stimulus-response (S-R) learning
basal ganglia function and, 575–577
fear in, 906–909
habits and, 906
Stimulus-stimulus (S-S), 906–909
STRAIGHT algorithm, in auditory object analysis, 370
Streaming, in auditory object analysis, 370–371
Stress hyporesponsive period (SHRP), 893
Striatum
effects of damage, 593–594
overview, 1094–1095
valuation in decision making, 1097
Strictly congruent, 628–629
Striosomal pathway, in basal ganglia function, 569
Stroke
aphasia following, 259–260, 810
Balint's syndrome following, 198
forebrain dysfunction following severe brain injury, 1129
impact of, 144
language processing after, 1238
neurodegenerative disease resulting from, 132
recovery of hand motor function, 148–149
white matter structural changes following, 1128
Strong-loops hypothesis (Crick and Koch), 1170
Stroop task, 156, 185, 192–193, 199, 707, 928, 931, 963, 969, 1024, 1025
Structural magnetic resonance imaging (MRI), of age-related change in cortical thickness, 73–74
Structural neuroimaging, comparative, 55–58
Structural organization of brain, neuronal circuitry of frontal lobe, 29–34, 36, 37, 39, 41
Subjective reports, in measuring visual awareness, 1151
Subpial granular layer (SG), cross-species comparisons of cortical development, 11, 22
Subplate in formation (SPF), neuronal circuitry of frontal lobe, 31–34
Subplate intermediate zone (SP/IZ), cross-species comparisons of cortical development, 11
Subplate zone (SP)
cross-species comparisons of cortical development, 10, 11, 13, 18, 21
neuronal circuitry of frontal lobe, 35–36, 37, 38, 39–40, 43
Subplate zone (SZ), neuronal circuitry of frontal lobe, 30, 31, 33, 35
Subsidiarity, 1021
Substantia nigra (SN)
basal ganglia function, 569–571
virus tracing studies, 553–558, 567–568
Subthalamic nucleus (STN), basal ganglia function, 569–571
Subtraction paradigm, 1257
Subventricular zone (SVZ)
cross-species comparisons of cortical development, 9, 11, 14, 20, 21, 22
neuronal circuitry of frontal lobe, 31, 32–33, 34–35, 36
Superior frontal gyrus (SFG), mirror neural system and, 634
Superior frontal sulcus (SFS), in visuospatial working memory, 74
Superior longitudinal fasciculus (SLF), dense perisylvian white matter connectivity, 260–264
Superior/middle temporal gyrus (STG/MTG)
dense perisylvian white matter connectivity, 260, 263
grammar processing and, 1236–1237
in phonological processing
speech production, 770
spoken word recognition, 768–769
in semantic unification network, 828–832
Superior occipitofrontal fasciculus (SOF), dense perisylvian white matter connectivity, 261, 264
Superior temporal sulcus (STS)
grammar processing and, 1236–1237
moral judgment and, 988, 991
object concepts and, 1034–1040
in phonological processing
critical role, 769
spoken word recognition, 768–769
working memory in auditory-motor integration networks, 769
in semantic unification network, 828
Supplementary motor area (SMA), 142, 143, 555–556
action understanding, 647
anatomical versus representational hierarchy, 643–644
in decision making, 1025–1026
goal representation in bimanual coordination, 646
spontaneous motor initiation, 1191
Surface perception, 435–449
active interpolation theory
described, 435
evidence for, 439–442
features of early visual system, 437–439
adaptation, defined, 436–437
contour in, 133–134, 135, 136
experimental paradigms for, 436
filling-in research
assumptions and insights, 447–449
computational model for, 442–447
definition of perceptual filling-in, 436–437
future directions, 449
limitations, 449
simulation studies, 443–447
problem of, 435–437
simulation studies, 443–447
texture in, 135, 153, 443–447
model for texture filling-in, 442–447
theoretical background, 435–437
Surface-related spreading activation, 438
Surface segmentation, 131
Surprise, facial expression and amygdala response, 936–939
Syllable discrimination, in phonological processing, 768
Sylvian fissure, 186, 263–264
Symbolic encoding theory, 436
Symmetry of perception, cortical, 316–318
Synaptic connections
cross-species comparisons of neocortical development, 7, 19–20, 22
neuronal circuitry of frontal lobe, 35
plasticity of. *See* Synaptic plasticity
Synaptic plasticity, 109–123
attractor dynamics in neural networks, 121–123
complementary mechanisms in control of, 146–147
long-term depression (LTD) of cellular mechanisms, 109–123
long-term potentiation (LTP) of cellular mechanisms, 109–123
modifications as means for memory, 109–111
Synaptic tagging, 114

Synaptogenesis
cross-species comparisons of cortical development, 22
neuronal circuitry of frontal lobe, 35, 37, 40–41, 42
Synchronization of neurons. *See* Selective attention, through neuronal synchronization
Synergies. *See* Muscle synergies
Syntactic processing, 169
deficits in, 807–814
lesion studies, 807–813
mediating role of syntax, 880
neuroimaging studies, 810–813, 845
syntactic representations and processing, 805–807, 845
syntactic structures, 805–806
syntax-semantics framework, 824–827
System identification, in motor learning and control, 588

T

Tabula rasa hypothesis, in cross-species comparisons of cortical development, 18
Task dependence, in phonological processing, 767–768
Task difficulty, in learning and learning transfer, 159–160
Task sets, working memory and, 228–229
Task-specific deactivations, 1068
Task-switching, as cognitive control function, 76
Telencephalic pallium, neuronal circuitry of frontal lobe, 30
Telencephalon, cross-species comparisons of neocortical development, 8–9, 10
Temporal decoding, of neural population responses, 428–429
Temporal difference model, in valuation process, 1087–1088
Temporal encoding
in odor discrimination, 329–330
in perceptual learning, 135–136
Temporal encoding function (TEF), 604–605
Temporal expectancies, selective neuronal synchronization and, 296–297
Temporal tasks, auditory learning in, 358–359
relative-timing tasks, 359, 360
temporal-interval discrimination, 358–359
Temporal word form area, 1153
Temporoparietal junction (TPJ)
moral judgment and, 988–989, 991, 993–994
perisylvian network for spatial orienting, 259–260
social cognition and, 955
Tetris (video game), 159
Texture determination, 135, 153
computational model for texture filling-in, 442–447
simulation studies in, 443–447
Thalamic neurons
central thalamic deep-brain stimulation, 1131–1134
in forebrain dysfunction following severe brain injury, 1128–1130
Thalamic radiation (TR), cross-species comparisons of cortical development, 13
Thalamocortical fibers, neuronal circuitry of frontal lobe, 32–33, 37–39, 43
Thalamus
in cognitive function, 565, 567
neuronal circuitry of frontal lobe, 31, 32, 35–36
receptive fields in primary visual cortex, 409–411
transfer of information to auditory cortex, 95
transfer of information to visual cortex, 91, 92
THBS4 (thrombospondin 4), 60
Theories of competence, 1241
Theories of processing, 1241
Theory of mind (TOM), 41, 758, 880, 954, 1042
Thorndike, E. L., 739
Thought experiments, 1115, 1202–1203, 1211–1212
Threat detection, 957–958
Three-dimensional structural coding hypothesis, 460–463
Three-dimensional surface orientation, 488–489
Tics, 572
Time
in perceptual learning, 135–136
synaptic plasticity and, 122–123
Time-dependent behavioral impairment, 694–695
Tinnitus, 147
Todorov, E., 549
Token Test, 809
Tonically active neurons (TANs), basal ganglia function and, 574–575
Top-down influences. *See also* Cognitive (executive) control
in consciousness, 1195–1197
in learning process, 134, 136–137
in selective attention, 186–187, 211–212, 213–214, 219, 220
causality of biases, 223, 224, 230
physiological mechanisms for, 229–230
relation to bottom-up influences, 213–214
selective neuronal synchronization, 289–299
Topic maps, 1230–1233
Topographic projections/maps
of dorsal attention network, 223–226
inherited color vision deficiencies, 388–392
retinotopic, 91–102, 223–226
Tourette syndrome, 82
basal ganglia function and, 568, 571, 572
TRACE, 768
Trace deletion hypothesis, 808
Training
attention training for children, 173–175, 176
auditory
characteristics of auditory processing in perceptual-learning patterns, 358–362
generalizations, 355–357, 360
neural processes related to auditory tasks, 353–358
inhibitory avoidance, 116
parent, 176
plasticity related to, 153–161
complex learning environments, 156–158
determinants of learning and learning transfer, 158–161
impact of practice, 154–156
specificity of learning, 153–154
training-related plasticity, 153–161
complex learning environments, 156–158
determinants of learning and learning transfer, 158–161
impact of practice, 154–156
specificity of learning, 153–154
Transcranial direct current stimulation (tDCS), 149
Transcranial magnetic stimulation (TMS)
of emotion regulation, 967, 970
empathy and, 975, 977, 982
functional role of perisylvian network, 263
goal representation and, 644–645
in methodology of cognitive neuroscience, 1256, 1259
posterior parietal cortex (PPC) function, 601, 603
role of attention in feature binding, 198
in semantic memory processing, 1062
spontaneous motor initiation, 1193–1194
vision pathologies, 1158–1159
Transient circuitry, neuronal circuitry of frontal lobe, 29, 32, 34, 38
Transient embryonic zones, cross-species comparisons of cortical development, 9–14
Transient form of paralysis, 1141
Tritanopia, 388–389
Tritiated thymidine, as marker for DNA synthesis, 8
Trolley problem, 991, 994
Troxler fading paradigm, 439, 440–441
Tryptophan hydroxylase-2 (TPH2), in imaging genetics studies, 948
Tulving, Endel, 655, 743, 752
Twin studies
genetic basis of emotional variability, 946
language impairment, 856, 863
Typically developing (TD) children, 171, 172, 174

U

Ultimatum Game (UG), 966–967, 994
Uncertainty
interactions between control and memory, 713–714
reduced, 715–719
Unconditioned stimulus (US), in decision making, 1095–1096

Unconscious vision, 1152–1155
Unification, in semantic unification framework, 819. *See also* Semantic processing, semantic unification
Unit recording experiments, 305–306
Useful Field of View Task, 157–158
Utilitarian judgment, moral judgment and, 991–992

V

Variance maps, 744, 745
Vegetative state (VS), 1124
Ventral frontoparietal attention network. *See* Frontoparietal attention network, ventral
Ventral intraparietal area (VIP), optic flow in primates and, 502–503
Ventral premotor area (PMv), 555–556
Ventral visual pathway, 789–794, 798–799
  reading process and
    interfacing with verbal system, 798–799
    pathology, 792, 794
  visual awareness and, 1159–1160
Ventricular zone (VZ)
  cross-species comparisons of cortical development, 9, 11, 13, 14, 16, 17, 18, 20, 21, 22
  neuronal circuitry of frontal lobe, 31, 33, 34–35, 36
Ventrolateral prefrontal cortex (VLPFC)
  development of, 74, 79, 81, 705, 706, 707, 709–714, 714
  in emotion regulation, 963
  memory and, 705, 706, 707, 709–720, 714, 932
  in processing of emotional information, 932
Ventromedial prefrontal cortex (VMPFC), moral judgment and, 987–990, 992, 994
Vernier acuity tasks, 159
Vernier discrimination, 136–137
Vestibular signals
  perception of heading from, 500–502
  primate cortical neuron response to, 502–503
Video games
  aggressive behavior and, 155
  in complex learning environments, 157–158
  dopamine release and, 161
  generalized learning and, 160
  mental rotation and, 157, 159
Violence, moral judgment and, 991
Virtual brain damage, 1005
Virus tracing studies, 554–561, 567–568
Visual analogy problems, 80
Visual attention. *See also* Visual awareness
  attentional control processes, 251–257
    brain networks for conflict processing, 253–255
    conflict and, 252–253
  consciousness versus, 199–200, 1141–1143
  deficits in, 269–277
    complete, 270–273
    loss of perceptual space, 269–270
    partial, 273–277
  effect on visual neurons, 281–286
    attentional bottleneck, 282–283
    attentional modulation of neuronal responses, 285–286
    selectivity of visual neurons, 284–285
    strength of sensory responses, 283–284
  feedback in, 1165–1175
    anatomical observations, 1165–1166
    binocular rivalry role in, 1170–1176
    physiological observations, 1166–1168
    role of feedback in attention, 1168–1170
    visual masking role, 1170–1176
  flash suppression, 1142, 1144, 1145, 1154
  frontoparietal attention network and, 219–230
    anticipatory signals, 219–223
    coding locus of attention, 224–226
    eye movements and feature-based selection, 226–227
    functional connectivity, 219–223
    prefrontal-cingular circuits, 228–229
    reward/value signals and limbic system, 227–228
    task sets, 228–229
    top-down biases, 223
    topographic organization of maps, 223–224
    working memory, 228–229
  neural processing of emotional information, 925–926
  in reading process
    parts of words, 797–798
    pathology, 797
    serial decoding, 797–798
  selective, 157–158, 205–214. *See also* Selective attention
Visual awareness, 1151–1161
  brain activity and, 1152
  characterizing unconscious homunculus, 1152–1154
  empirical ad theoretical integration, 1159–1160
  feedback in, 1165–1175
    anatomical observations, 1165–1166
    binocular rivalry role in, 1170–1176
    physiological observations, 1166–1168
    role of feedback in attention, 1168–1170
    visual masking role, 1170–1176
  impact of, 1155–1158
  measuring, 1151–1152
  pathology, 1158–1159
  unconscious vision and multivariate pattern analysis, 1154–1155
Visual cortex. *See also* Visual neurons; Visual processing
  anatomy of feedback in, 1165–1166
  attention and. *See also* Visual awareness
    spatial attention, 206–208
  comparative neuroimaging, 57–58
  consciousness and
    biological theory, 1112–1113
    visual awareness, 1155–1158
  cross-species comparisons of cortical development, 10, 19–20, 55, 57–58
  feedback in, physiology, 1167–1168
  formation of eye-specific connections, 67–71
    binocular vision, 68, 70, 99–102, 130–131
    inputs to dLGN, 68–70, 94
    ocular dominance columns, 67–68
    role of retinal waves, 68, 70–71
  grandmother cells, 309–319
    correlation, 316–318
    cortical function, 314–316
    defined, 310
    history, 309–314
    invariance, 316–318
    symmetry, 316–318
  hierarchical function, 314–316
  impact of visual deprivation on auditory development, 167–169
  object recognition, 455–465
    discriminative learning, 133–134
    network processing of boundary fragments, 463–464
    object classification, 159
    retinal signals, 455
    three-dimensional surface fragments, 460–463
    two-dimensional object boundaries, 455–460
  in perceptual learning, 132
  plasticity, 129–137, 166–167
    lesions and, 131–132
    neuronal activity in formation of eye-specific connections, 67–71
    ocular dominance plasticity, 91, 99–102
    perceptual learning and, 132–136
    postnatal development, 129–131
    top-down influences, 134, 136–137
  statistical connectivity theory, 411–415
    cortical maps and, 413–414
    described, 411–413
    testing, 414–415
  subvoxel specialization of, 448
  visual abilities and video games, 157
  visual pathway, 91–102
    activity-dependent refinement of visual maps, 98–99
    early development, 91–95
    ocular dominance plasticity, 99–102
    retinotopic projections/maps, 91–102
    rewiring vision into auditory pathway, 95–98, 131, 166–167
    in word reading, 789–800
  wiring of functional maps, 409–411
  wiring of receptive fields, 409–411
Visual extinction, 1159
Visual information, 321
Visual masking, 1170–1176
Visual neglect, 1159
Visual neurons. *See also* Visual awareness; Visual cortex; Visual processing
  effect of visual attention on, 281–286. *See also* Visual awareness
    attentional bottleneck, 282–283
    attentional modulation of neuronal responses, 285–286

Visual neurons. (continued)
selectivity of visual neurons, 284–285
strength of sensory responses, 283–284
timing of neural activity, 1155
selective synchronization of, 289–299
attention selection in neuronal communication structure, 290–292
selective attention through selective synchronization, 292
selective inter-areal synchronization, 298–299
selective modulation of synchronization, 293–298
synchronization in interneuron networks, 292–293
Visual processing. *See also* Color vision; Surface perception
learning versus perception, 675
memory impairment and, 675–678
for perceiving objects, 1032–1042
Visual representations. *See also* Object recognition
phonological impact on, 799
word perception as object perception, 789–792
Visual stimuli, speech-related, 771
Visual Word Form Area (VWFA), 792–795, 798, 799, 800
Visuomotor coordination task, 149
Visuospatial working memory, development of, 74–75
Vocabulary development
early visual processing of printed words, 791–794
neural signatures of word learning, 842–843
spoken word recognition, 768–769
Vocal learning. *See also* Speech production; Speech recognition
mechanisms underlying, 878
Volition. *See* Consciousness, volition and function of
Volleyball, complex learning environments, 156–157
Voltage-gated (N-type) receptors/channels, cross-species comparisons of cortical development, 15
Voltage-sensitive dye imaging (VSDI), visual cortex, 419–423, 425–432
Voluntary actions. *See* Consciousness
Voluntary motor control. *See* Motor processing
Von Economo neurons, in human brain specialization, 54–55
Voronoi analysis, 388–389
VWFA (Visual Word Form Area), 792–795, 798, 799, 800
Vygotsky, Lev, 161

W

Wada procedures, word recognition and, 768
Wallis, Terry, 1127–1128
Water polo, complex learning environments, 156–157
Wechsler Intelligence Scale for Children (WISC-III), 156
Wernicke, Carl, 1235
Wernicke-Geschwind model, 741
Wernicke's area
comparative neuroimaging, 57
connectivity studies, 261
in human brain specialization, 53, 54
spoken word recognition and, 768
Wernicke's language model, 773
Western Aphasia Battery, 809
White matter (WM). *See also* Gray matter
in cognitive control development, 76
comparative neuroimaging, 55–58
neuronal circuitry of frontal lobe, 31, 37
perisylvian white matter connectivity, 260–262
and plasticity of human neurocognition, 165
structural changes following brain injury, 1127–1128
Whitening, 427, 428
Wiesel, Torsten, 309, 311–312, 314, 409
Williams syndrome, 166
Wilson, Eric, 961
Win-shift behavior, 575
Word reading. *See* Reading process
Working memory
auditory-motor integration networks in, 771
consolidation theory and, 692–695
in decision making, 1021
defined, 678
development of, 74–76, 82
anatomical perspectives, 566
manipulation of information in, 75–76
nonspatial working memory, 75
visuospatial working memory, 74–75
lesion studies of medial temporal lobe function, 678–679
long-term memory versus, 683
nature of, 73
path integration and, 683–684
in visual selection process, neuroimaging studies of, 228–229
World knowledge, in semantic unification framework, 822
Writing, striatal damage and, 593–594

X

X-linked recessive traits, 388–389, 390

Y

Yerkes-Dodson law, 161

Z

Zombie behaviors, 1145–1146, 1155
Zone of proximal development (Vygotsky), 161

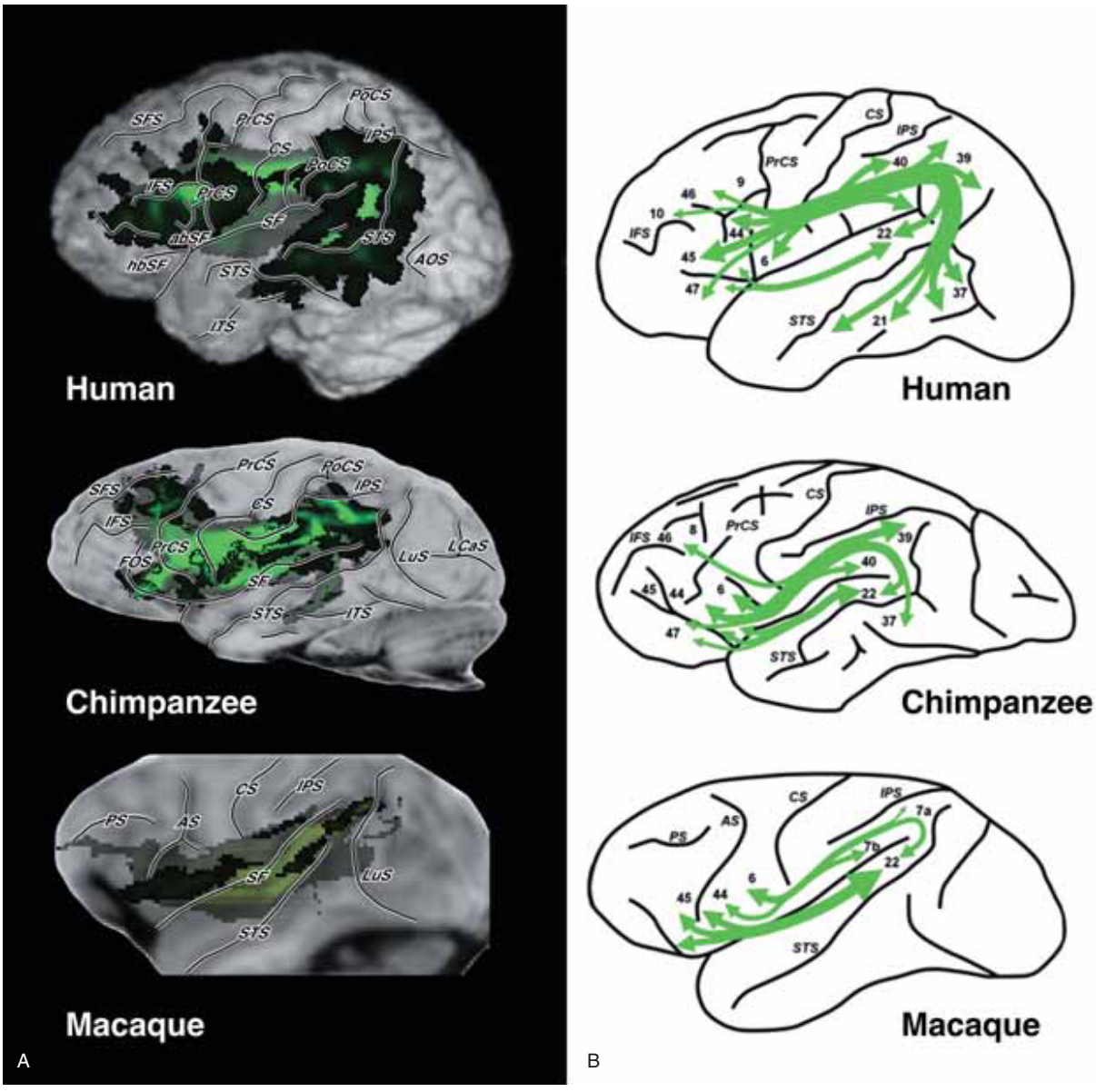


PLATE 1 Evolution of the human arcuate fasciculus (AF), which interconnects frontal and temporal language areas, based on the comparative diffusion-tensor imaging (DTI) results of Rilling and colleagues (2008). (*A*) Average tractography results from the left hemispheres from 10 humans, three chimpanzees, and two macaque monkeys. (*B*) Schematic representation of the results shown in *A*, representing the cortical endpoints of the tracts in terms of Brodmann's areas. Both humans and chimpanzees have a distinct AF, although in humans the AF includes strong connections with the temporal cortex below the superior temporal sulcus (STS), including area 21, a region where word meaning is represented. Chimpanzees have very few fibers in the AF that extend to the cortex inferior to the STS. Macaques do not have a definite AF: fibers traveling between the posterior inferior frontal cortex and posterior temporal lobe take a more ventral route, passing deep to the insula, and include few if any fibers with endpoints inferior to the STS. (See figure 3.4.)

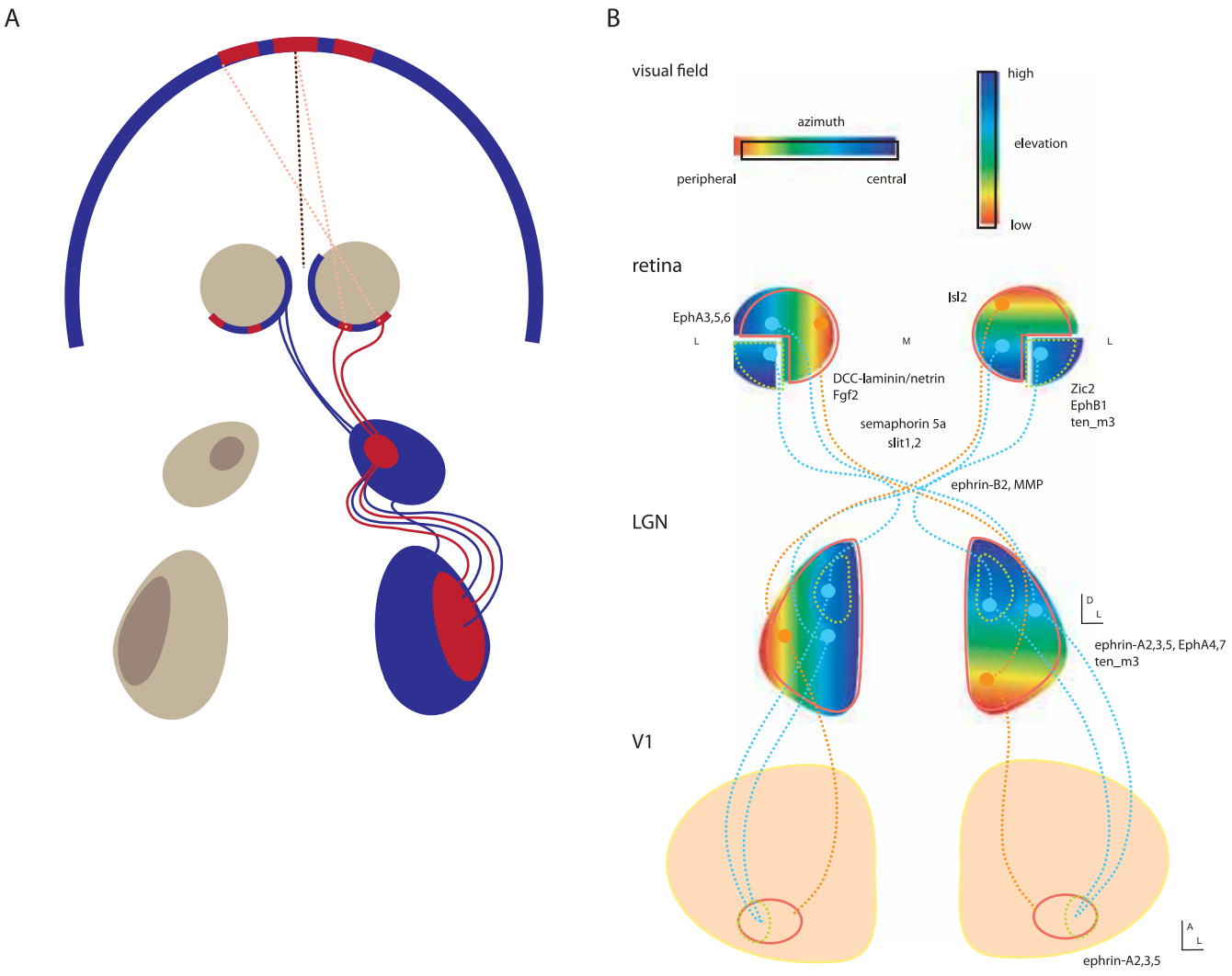


PLATE 2 (*A*) Representation of the rodent visual pathway. Retinal ganglion cells project to the LGN, which in turn projects to the primary visual cortex (V1). A central region of the visual field is represented by both eyes along the pathway (ipsilateral, red; contralateral, blue). Contralateral and ipsilateral retinal ganglion cell terminals representing this binocular region are segregated in the LGN (red, ipsilateral zone; blue, contralateral zone). Geniculocortical fibers representing this region converge onto a binocular zone located in the lateral half of V1 (red, binocular zone; blue, monocular zone). (*B*) Schematic representation illustrating retinotopic map organization at each stage of the visual pathway and known guidance cues contributing to patterning. The visual field can be divided into two Cartesian axes, azimuth and elevation. For clarity, the azimuthal map on the left is diagrammed onto the visual pathway of the right hemisphere. The elevation map on the right is diagrammed onto the pathway of the left hemisphere. In reality, both axes of visual space are represented concurrently in both hemispheres. The ganglion cell sheet of the retina is divided into a contralaterally projecting region and an ipsilaterally projecting region. The ipsilateral retina originates from the ventrotemporal quadrant and is characterized in late embryogenesis by Zic2 and EphB1 expression. Conversely, the contralateral retina is characterized by Isl2 expression. Retinal ganglion cells express DCC and are repulsed out of the optic head by laminin and netrin. Factors, such as semaphoring-5a, keep retinal axons on course in the optic tract, where ipsilateral axons are repulsed by ephrin-B2 while contralateral axons decussate. High temporal to low nasal gradients of EphA receptor and ten_m3 expression in retinal axons likely influence terminal zones onto gradients of ephrin-A in the LGN. Ipsilateral axons terminate in a dorsomedial core of the LGN, segregated from surrounding contralateral axons. Activity-dependent refinement is necessary for proper eye-specific segregation. While ephrin-A gradients shape retinotopic termination zones, ten_m3 specifically influences ipsilateral targeting. Geniculocortical axons innervate V1. Ipsilateral inputs and corresponding contralateral fibers converge in the lateral binocular zone, while contralateral inputs representing regions not detected by the ipsilateral eye terminate in the medial monocular zone. Loss of ephrin-As leads to the disorganization of cortical maps only on the azimuthal axis, suggesting that other, unidentified factors contribute to the mapping of elevation. (See figure 6.1.)

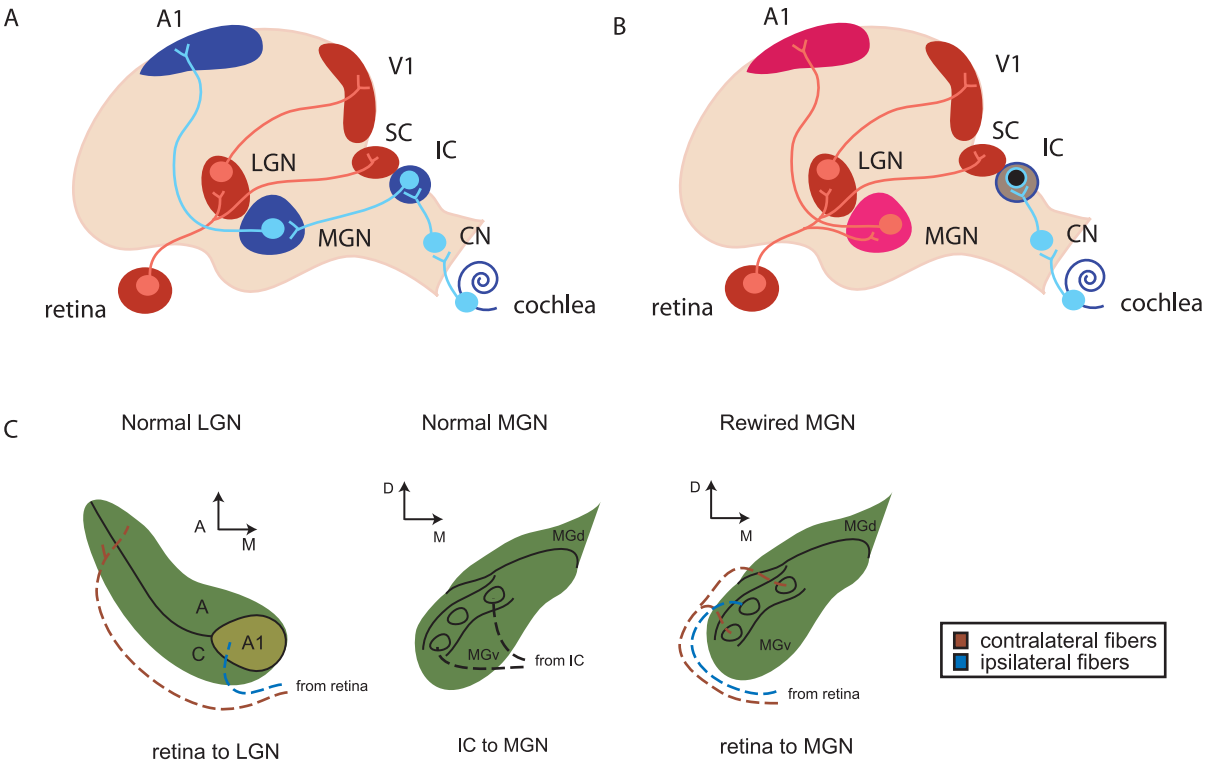
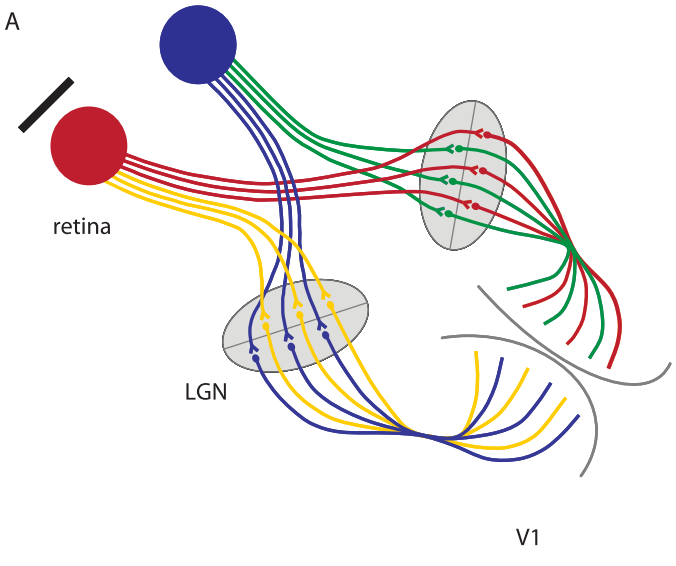
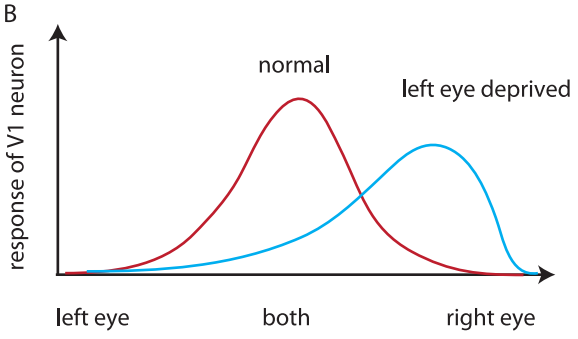


PLATE 3 Primary visual and auditory pathways in normal and rewired mice: anatomical and physiological consequences of rewiring. (*A*) The visual pathway in ferrets and mice begins with retinal projections to the lateral geniculate nucleus (LGN) and superior colliculus (SC). The LGN projects in turn to the primary visual cortex (V1). The auditory pathway traces from the cochlea to the cochlear nucleus (CN) and then to the inferior colliculus (IC). From IC, connections are made with the medial geniculate nucleus (MGN), which projects to the primary auditory cortex (A1). (*B*) Ablation of the IC in neonatal animals induces retinal afferents to innervate the MGN and drive the auditory cortex to process visual information. (*C*) Retinogeniculate axons of normal ferrets project to eye-specific regions of the LGN (horizontal plane), while IC afferents project to the ventral subdivision (MGv) of the MGN (coronal plane) and innervate lamellae parallel to the lateral-medial axis. Rewired auditory fibers innervate the MGv along adjacent, nonoverlapping eye-specific terminals within MGv lamellae. (Adapted from Sur & Leamey, 2001.) (*D*) Orientation maps are present in normal V1 and rewired A1 of ferrets using optical imaging of intrinsic signals. The animal is stimulated with gratings of different orientations, while hemodynamic changes in red wavelength light reflectance caused by increases in oxygen consumption are detected from the cortex with a digital camera. The orientation preference map is calculated by computing a vector average of the response signal at each pixel. Color bar: color coding representing different orientations. Scale bar: 0.5 mm. (*E*) Retrograde tracers reveal the pattern of horizontal connections in superficial layers of normal V1, normal A1, and rewired A1 of ferrets. Distribution of horizontal connections in rewired A1 more closely resembles that of normal V1 than normal A1 and potentially contributes to the refinement of orientation mapping in rewired A1. Scale bars: 500 μm. (Adapted from Sharma, Angelucci, & Sur, 2000.) (See figure 6.2.)





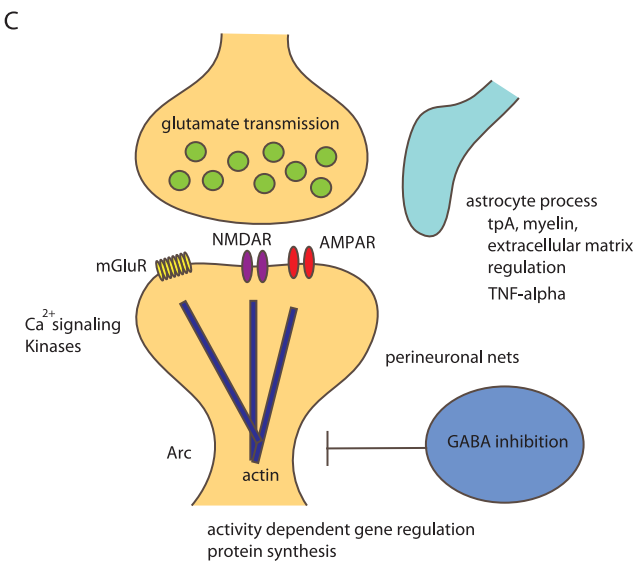


PLATE 4 Ocular dominance anatomy and plasticity in V1. (*A*) Contralateral and ipsilateral fibers are segregated in the LGN but converge onto binocular cells in V1. (*B*) When one eye is deprived of input for a brief period during the critical period, or for a longer period during adulthood, binocular cells in V1 become more strongly driven by the nondeprived eye. Ocular dominance plasticity reflects both structural and functional changes of synapses. (*C*) The cellular and molecular mechanisms of ocular dominance plasticity are an active area of investigation. Processes known to play a critical role include signal transduction pathways downstream of the mGluR and NMDARs, and activity-dependent changes in AMPAR content at synapses, mRNA transcription, and protein translation. GABAergic inhibition is involved in inducing the critical period of ocular dominance plasticity, and extracellular matrix factors and perineuronal nets surrounding inhibitory interneurons have been implicated in constraining plasticity. (See figure 6.3.)

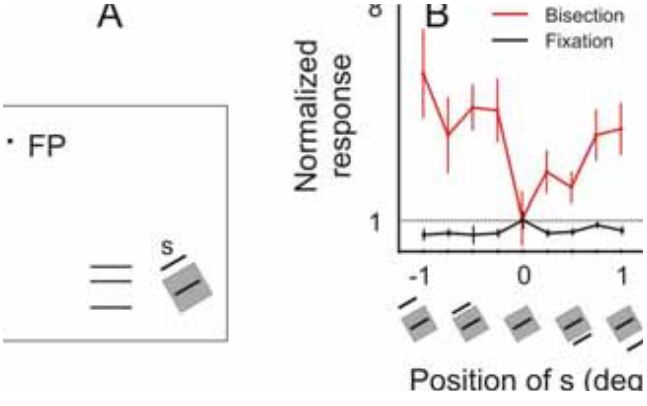


PLATE 5 Perceptual learning modifies contextual influences in V1. (*A*) The stimulus paradigm. The three horizontal parallel lines indicate the task stimulus. Monkeys were trained to determine whether the middle line was closer to the upper or the lower flanker. After training monkeys on this bisection discrimination task, responses of single V1 neurons to another stimulus, the test stimulus, were recorded when the animal either performed the trained bisection task, or simply maintained its fixation at the fixation point (FP). The test stimulus consisted of two lines, an optimally oriented line fixed in the center of the receptive field (denoted by the gray square), and a second parallel line (indicated by "s") placed at different locations on either side of the RF (see the cartoons at the bottom of *B*). (*B*) The normalized responses of a typical V1 cell to the test stimulus as a function of the position of line "s." When the animal was performing the simple fixation task, placing "s" on either side of the RF slightly suppressed neuronal responses relative to the responses at position 0 deg, where the two test lines were superimposed in the RF center. In contrast, when the animal was performing the bisection task, the weak contextual inhibition was changed into strong facilitation. (Adapted from Crist, Li, & Gilbert, 2001.)(See figure 8.3.)

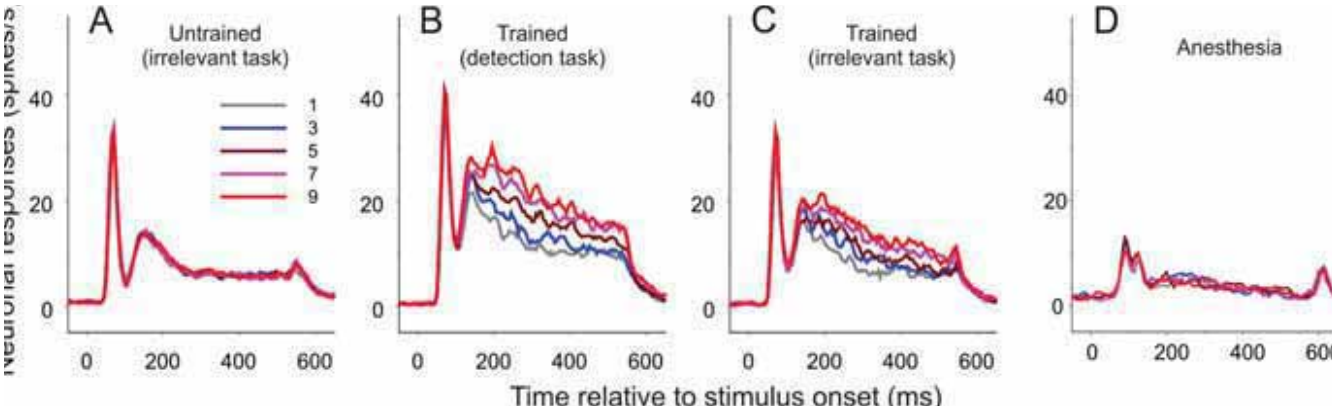


PLATE 6 Learning- and task-dependent changes in V1 associated with training on contour detection. Shown here are averaged population neuronal responses to visual contours consisting of 1, 3, 5, 7, and 9 collinear lines embedded in an array of randomly oriented lines (for example see figure 8.1). Time 0 indicates stimulus onset. (*A*) Neuronal responses in V1 of untrained monkeys are independent of contour lengths (the six peristimulus time histograms are superimposed), indicating the absence of contour information in V1 responses. (*B*) Over the course of training the animals on contour detection, a late response component associated with contour saliency emerges—the longer the contours, the stronger the neuronal responses. (*C*) In trained animals the contour-related V1 responses are much weakened when the animals perform tasks that are irrelevant to contour detection. (*D*) Contour-related responses disappear in the trained V1 region under anesthesia. (Adapted from Li, Piech, & Gilbert, 2008.) (See figure 8.4.)

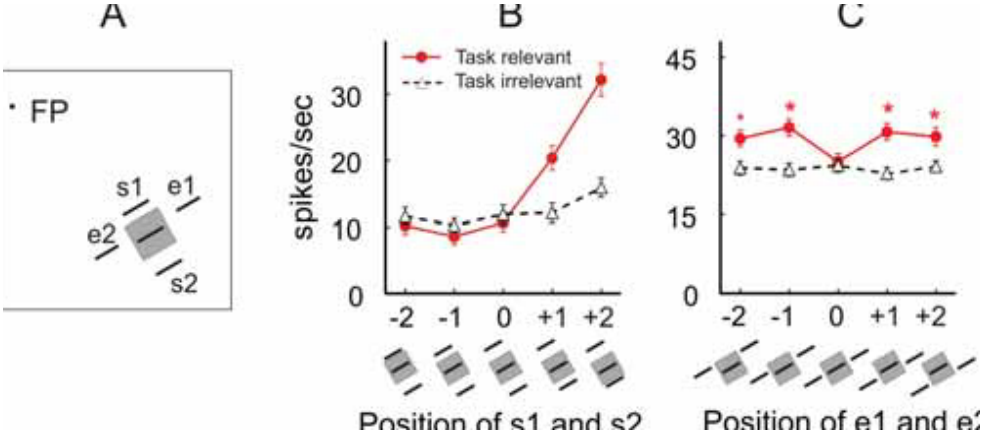


PLATE 7 Task-specific top-down influences on V1 responses. (*A*) Monkeys were trained to do two different discrimination tasks with identical stimulus patterns at the same visual field location. The stimuli consisted of five simultaneously presented lines: an optimally oriented line fixed in the RF center and flanked by four additional lines surrounding the RF. In different trials, the arrangement of the two side flankers (s1, s2) was randomly assigned from a set of five different configurations (illustrated in the cartoons at the bottom of *B*, labeled from −2 to +2). Each configuration differs from the others in the separation between the three side-by-side lines (in condition 0 the three lines were equidistant; in the other conditions either s1 or s2 was closer to the central line). In the same trials, the two end-flankers (e1, e2) were also independently assigned a random configuration from a set of predefined arrangements, such that the end flankers were collinear with each other but misaligned with the central line to either side (the cartoons at the bottom of *C*). The animal was cued to perform either a bisection task based on the three side-by-side lines or a vernier task based on the three end-to-end lines, using the same set of five-line stimuli. (*B*) Responses of a V1 cell were examined as a function of the position of the two side flankers s1 and s2 when the animal either performed the bisection task, in which s1 and s2 were task-relevant; or performed the vernier task, in which the same s1 and s2 were task-irrelevant. (*C*) Responses of a V1 cell were examined as a function of the position of the two end flankers e1 and e2 when the animal either performed the vernier task, in which e1 and e2 were task-relevant; or performed the bisection task, in which the same e1 and e2 were task-irrelevant (Adapted from W. Li, Piech, & Gilbert, 2004.) (See figure 8.5.)

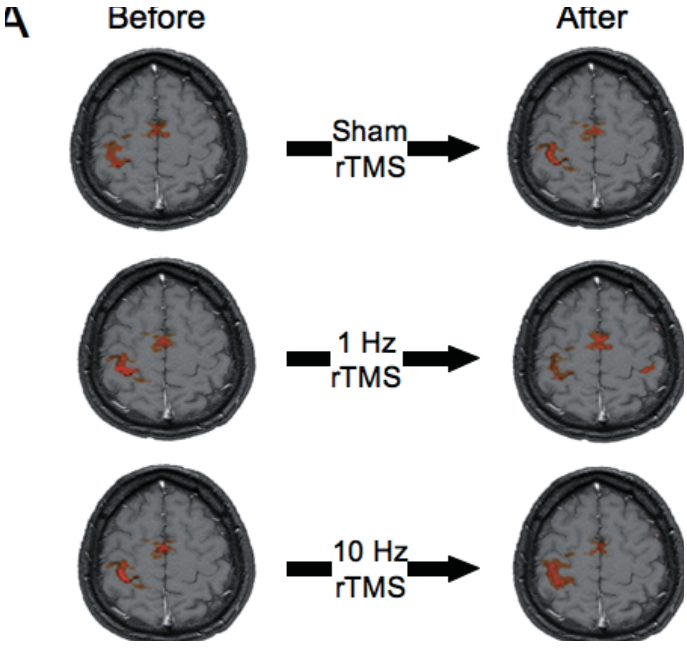


PLATE 8 (*A*) Brain activation in fMRI while subjects performed the same rhythmic hand movement (under careful kinematic control) before and after repetitive transcranial magnetic stimulation (rTMS) of the contralateral motor cortex. Following sham rTMS (*top row*) there is no change in the significant activation of the motor cortex (M1) contralateral to the moving hand and of the rostral supplementary motor cortex (SMA). After M1 activity is suppressed using 1-Hz rTMS (1,600 stimuli, 90% of motor threshold intensity; *middle row*), there is an increased activation of the rostral SMA and of M1 ipsilateral to the moving hand. Increasing excitability in the contralateral M1 using high-frequency rTMS (20 Hz, 90% of motor threshold intensity, 1,600 stimuli; *bottom row*) results in a decrease in activation of rostral SMA. (See figure 9.1A.)

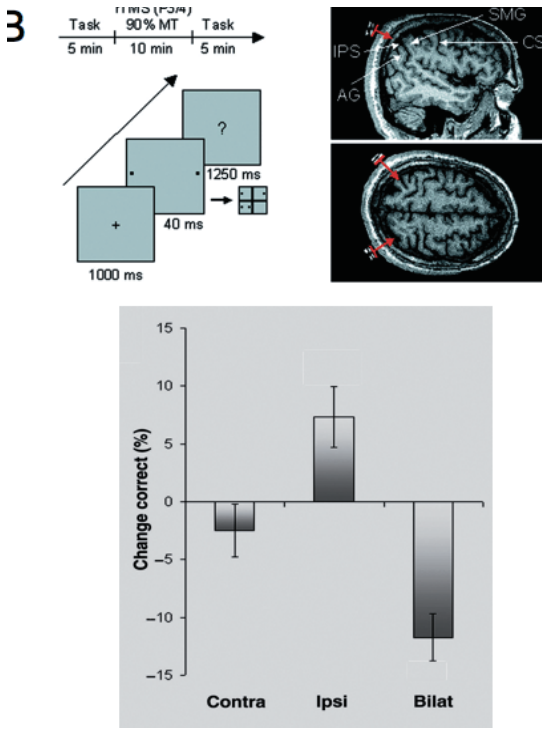


PLATE 9 (*B*) Impact of image-guided rTMS to the right parietal cortex on a visual stimulus detection task. During the task subjects were presented with carefully titrated visual stimuli on the right, left, or bilateral side of a computer monitor (*top left*) and had to respond by pressing the appropriate response button (right, left, or both). TMS was applied guided by the subject's own anatomical MRI using a frameless stereotaxic system (*top right*). There was a decrease in contralateral performance (neglect) but an even greater increase in performance ipsilateral to the parietal rTMS location (*bottom*). This summed up to a significant decrease in bilateral stimuli, where subjects neglected the contralateral stimulus and responded as if only the ipsilateral one had been presented (extinction of double simultaneous stimulation). (*B* modified from Hilgetag, Theoret, & Pascual-Leone, [2001]). (See figure 9.2B.)

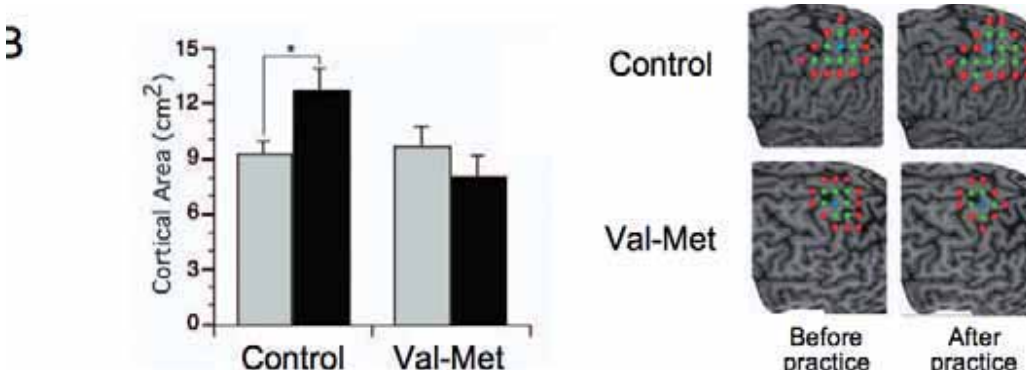


PLATE 10 (*B*) Histogram displaying the size of the cortical output maps before (gray bars) and after exercise (black bars) in control subjects and subjects with a val66met polymorphism for BDNF (*left side*). Following exercise, control subjects had significantly larger representations than at baseline, whereas subjects with a Met allele did not show a significant change. This difference is further illustrated by the representative motor maps from control and Val-Met polymorphism subjects superimposed onto a composite brain MRI image of the cortex (*right side*). Sites from which TMS evoked criterium responses in the target muscle are marked in green; negative sites are marked in red. (Modified from Kleim et al., 2006.) (See figure 9.3B.)

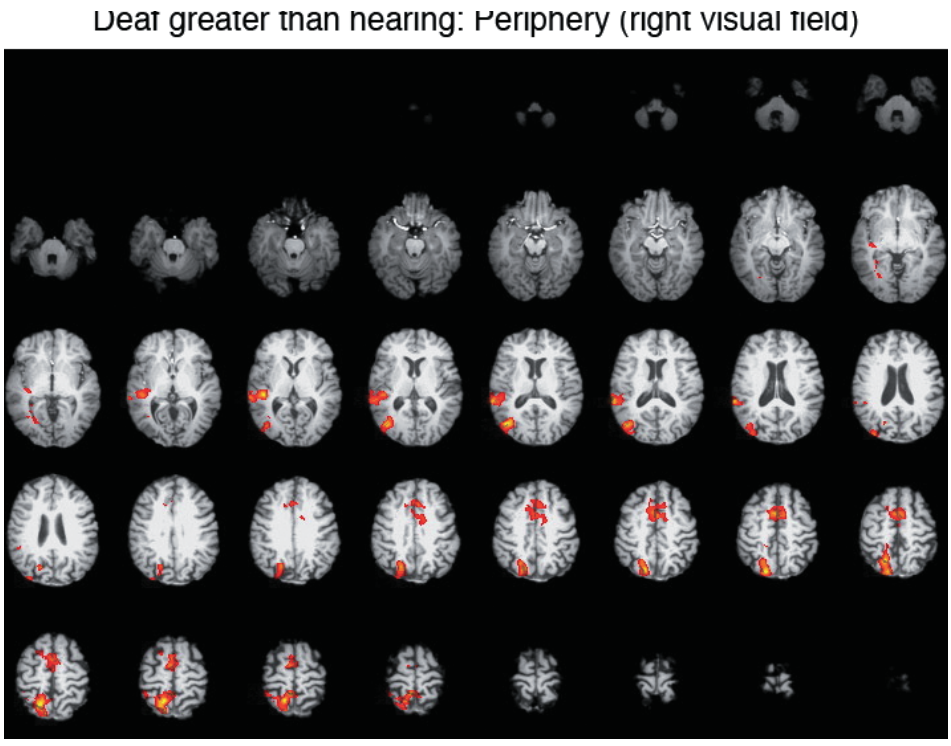


PLATE 11 Deaf and hearing participants completed a visual retinotopy experiment that included mapping of far peripheral visual space. The data show regions where activation was greater in deaf versus hearing participants in response to more peripheral visual stimuli presented in two distinct experiments (45–56° versus 11–23° and 11–15° versus 2–7°). Significant clusters included contralateral auditory cortex, STS, MT, anterior visual cortex, IPS, and anterior cingulate. (See figure 11.2.)

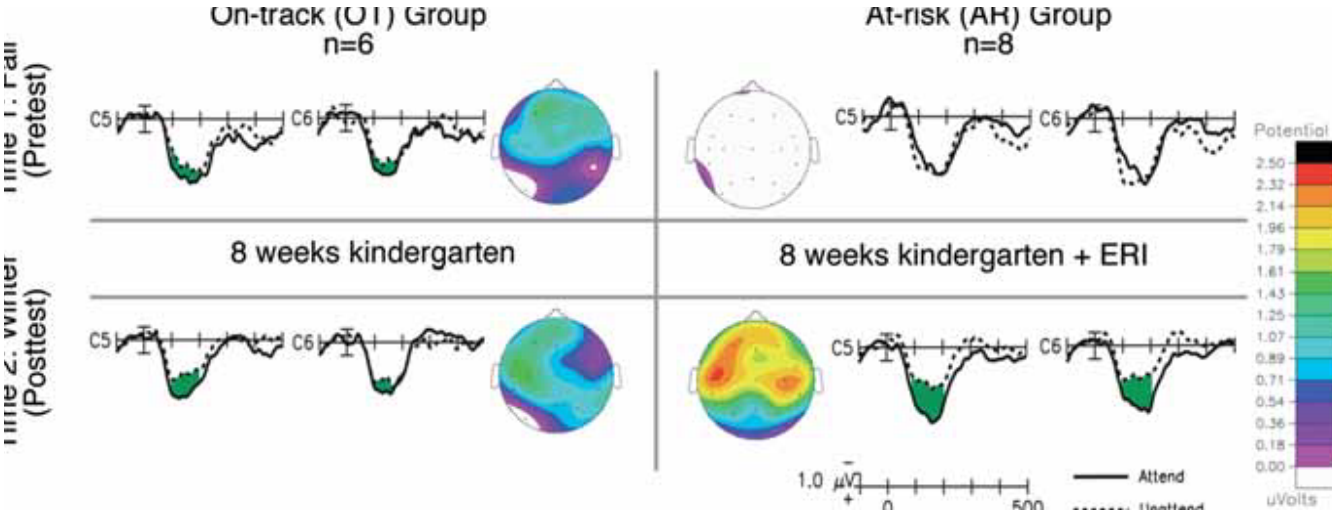


PLATE 12 Grand average ERP waveforms from the selective auditory attention paradigm show the effects of attention on sensorineural processing in kindergarten children of diverse early reading ability across the first semester of kindergarten. Top row shows data from pretest, and bottom row shows data from posttest for five-year-old kindergarten children on track (OT) in early literacy skills or at risk (AR) for reading difficulty. The OT group received eight weeks of kindergarten between pretest and posttest. The AR group received eight weeks of kindergarten with 45 minutes of daily, supplemental instruction with the Early Reading Intervention (ERI). Voltage map indicates the magnitude and distribution of the attention effect (Attended-Unattended). Changes in the effects of attention differed from pretest to posttest in the two groups ($P < .05$), with the OT group showing no change ($P = .92$) and the AR group showing a significant increase in the attention effect ($P < .01$). At pretest, the OT group tended to have a larger attention effect than the AR group ($P = .06$). At posttest, the AR group had a nonsignificantly larger attention effect than the OT group ($P = .17$). (See figure 11.7.)

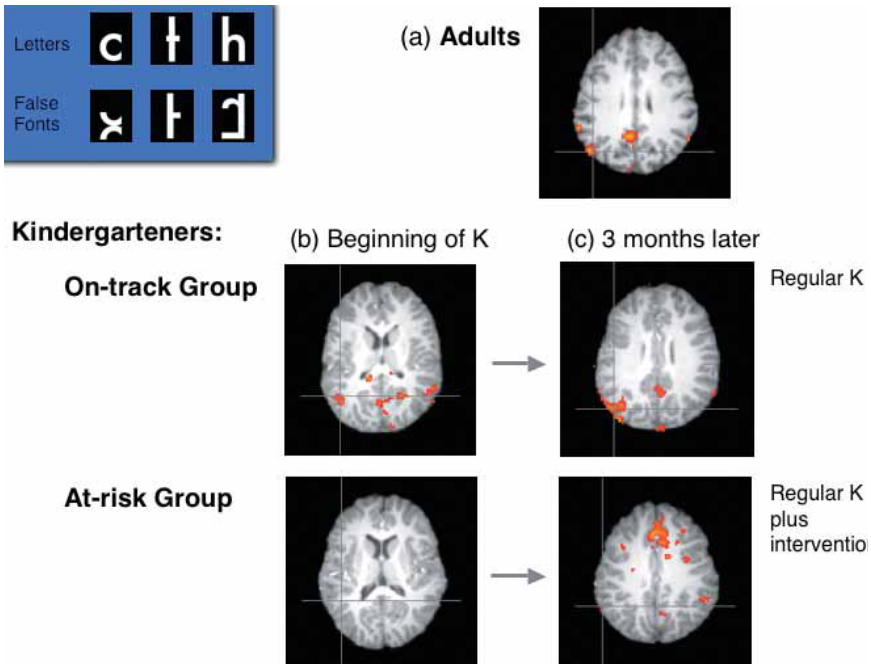


PLATE 13 Functional MRI activations for letter > false font while performing a 1-back task in adults and kindergarten children of diverse reading ability across the first semester of formal reading instruction. (*A*) Adults performing the task displayed activation in classic left temporoparietal regions. (*B*) In contrast, at the beginning of kindergarten, children on track in early literacy skills (upper panel) showed bilateral temporoparietal activation, and children at risk for reading difficulty (lower panel) showed no regions of greater activation. (*C*) Following one semester of kindergarten and, for children in the at-risk group, daily supplemental instruction with the Early Reading Intervention, on-track children showed left-lateralized activation in temporoparietal regions, and at-risk children showed bilateral temporoparietal activation and large activation of frontal regions, including the ACC. The left hemisphere is displayed on the left. In the upper left corner are example stimuli. (See figure 11.8.)

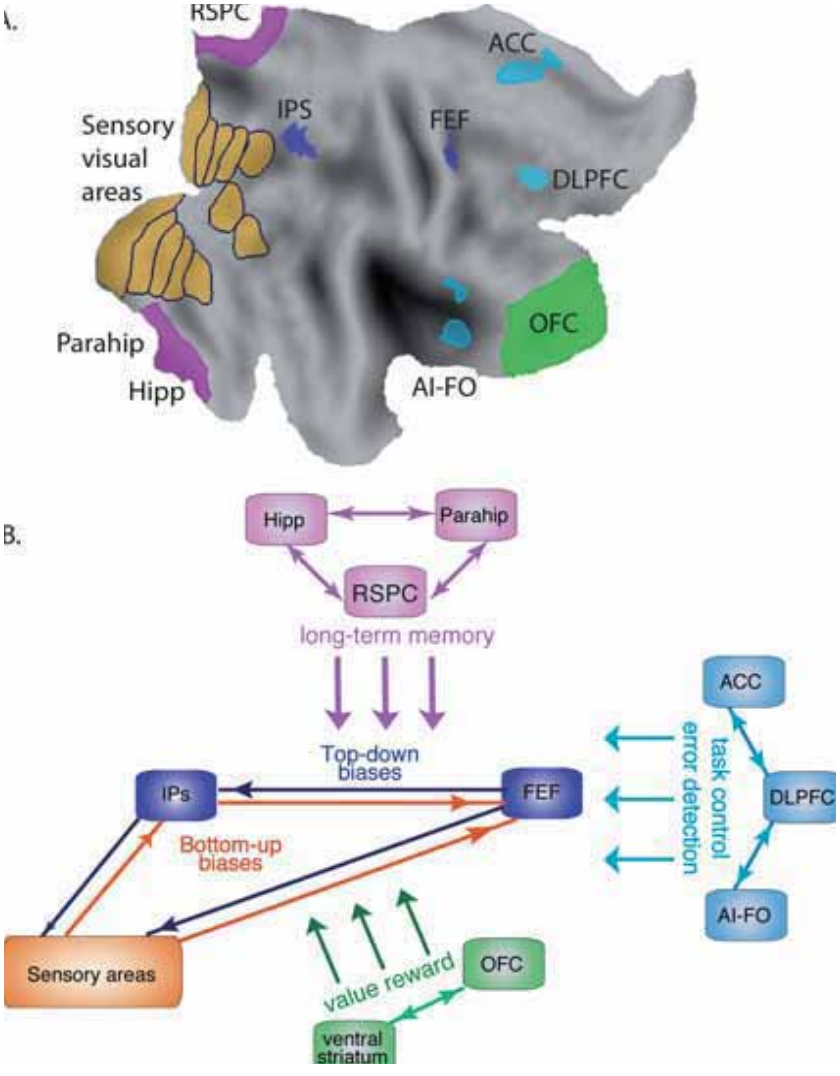


PLATE 14 (*A*) Control and data-processing networks. Flat map of right hemisphere on which regions and different networks involved in control are superimposed. Dark blue: dorsal frontoparietal attention network. IPS: intraparietal sulcus; FEF: frontal eye field. Orange: sensory visual areas. Purple: Long-term memory retrieval network. RSPC: retrosplenial cortex; Parahip: parahippocampus; Hipp: hippocampus. Azure: executive control network. ACC: anterior cingulate; DLPFC: dorsolateral prefrontal cortex; AI-FO: anterior insula–frontal operculum. Green: reward value network. OFC: orbitofrontal cortex; ventral striatum (not shown). (*B*) Wire diagram. Dorsal attention network feeds top-down and receives bottom-up biases to/from sensory cortices for stimulus and response selection. Other networks bias sensory processing via interaction with dorsal attention network. (See figure 14.1.)

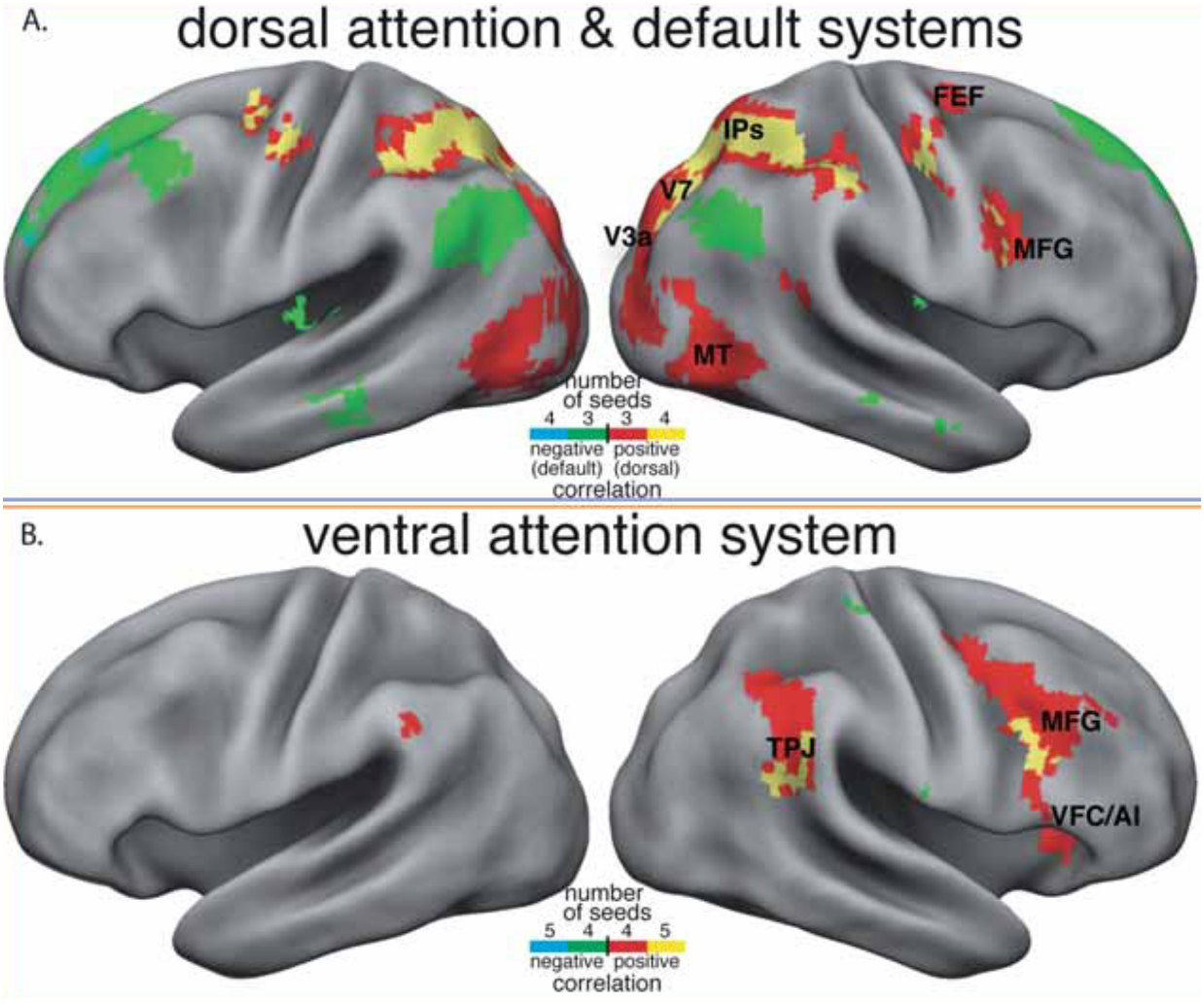


PLATE 15 Functional connectivity by fMRI (fcMRI) defines separate dorsal and ventral networks. (*A*) Dorsal attention and default networks. The map indicates regions that showed significant positive correlations with three (red) or four (yellow) of the seed regions in the dorsal attention network (IPS, FEF, V7, MT+). The dorsal network is largely reproduced in the resting state FC maps. Regions that show significant negative correlations with three (green) or four (blue) of the seed regions are also shown and roughly reproduce the default network, possibly indicating a push-pull relationship between the two networks. (*B*) Ventral attention network. Five ventral regions (R TPJ, R VFC, R MFG, R PrCe) were used as seeds for an FC analysis. Regions showing consistent positive correlations largely reproduce the ventral network, but negative correlations in default regions are not observed. The posterior MFG near the inferior frontal sulcus appears to be connected to both networks. (He et al., 2007.) (See figure 14.2.)

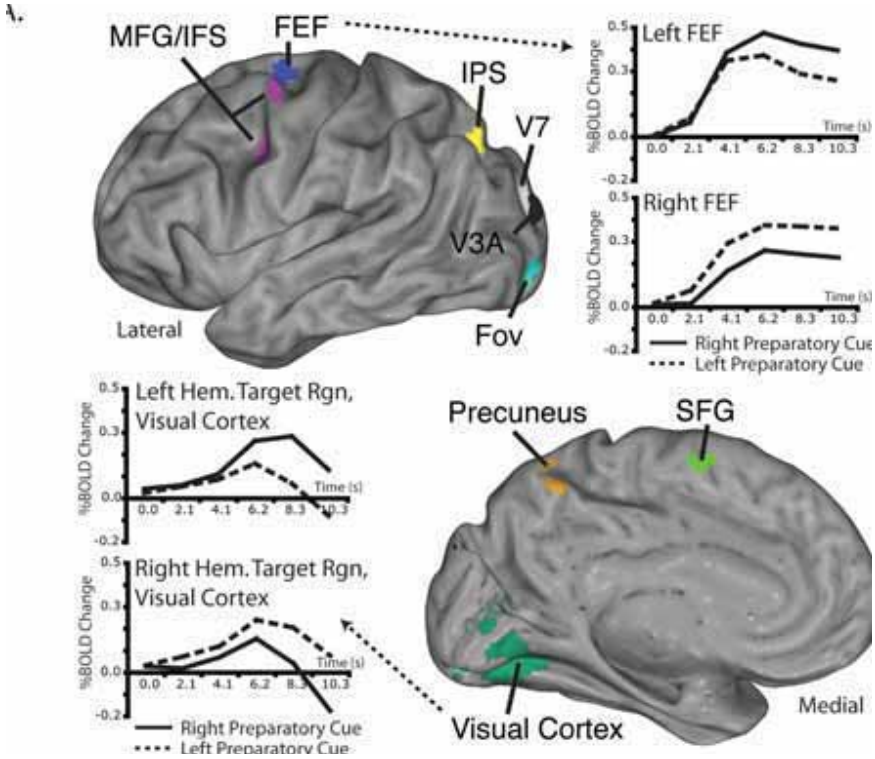


PLATE 16 (*A*) Frontoparietal areas and visual cortex modulated by anticipatory signal for spatial attention. Areas with spatially selective preparatory signals following an auditory cue directing attention to a left or right location. MFG/IFS: middle frontal gyrus/inferior frontal sulcus; FEF: frontal eye field; IPS: intraparietal sulcus; Fov: foveal region of V1–V3; SFG: superior frontal gyrus. BOLD signal time series following spatial auditory cues show anticipatory signals that are stronger for cue directing attention to contralateral visual field locations. (From Sylvester et al., 2007.) (See figure 14.3A.)

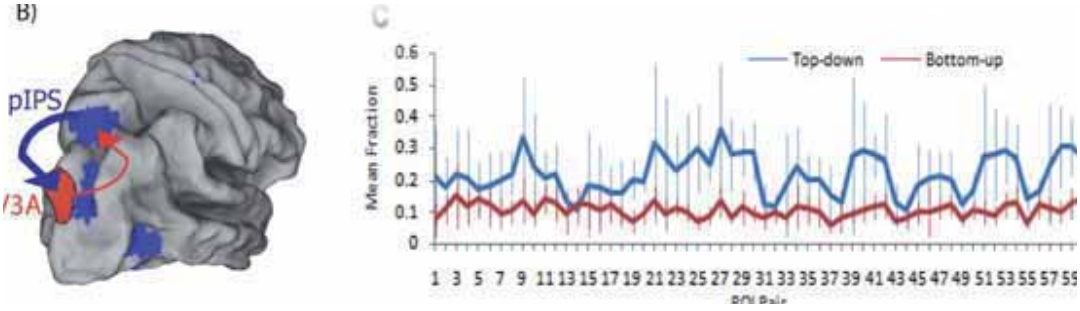


PLATE 17 (*B*) Granger causality of BOLD signal time series during anticipatory spatial attention. *Left*: "Control" area pIPS top-down modulates visual area V3A. The strength of top-down control trial-by-trial correlates with higher accuracy on a difficult visual discrimination task. *Right*: During anticipatory spatial attention the strength of top-down influences is stronger from control to visual areas than bottom-up influences from visual areas to control areas. (From Bressler, Tang, Sylvester, Shulman, & Corbetta, 2008.) (See figure 14.4B.)

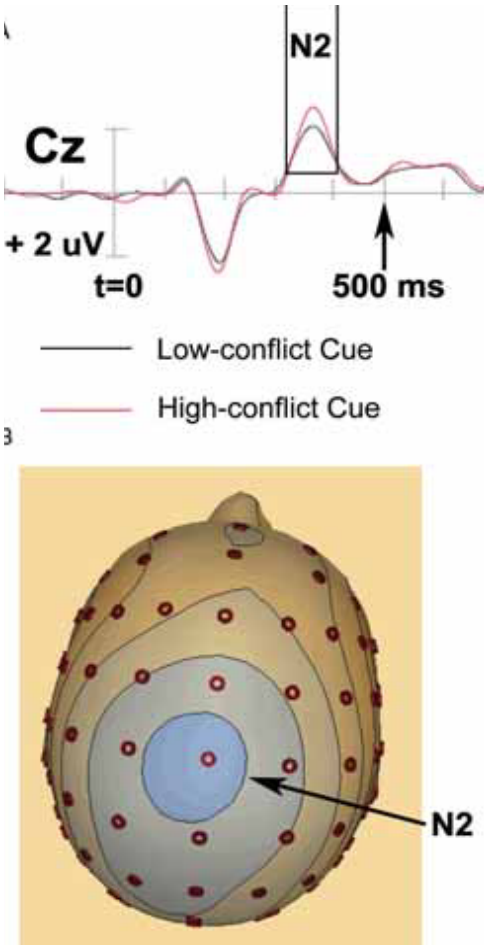


PLATE 18 (*A*) Grand average ERPs to cues as a function of conflict, recorded from midline central scalp electrode site Cz. Negative voltage is plotted upward, and cue onset is the upright bar at time zero. High-conflict cue ERPs are plotted in red. The box indicates the N2 component, peaking at approximately 360 ms from cue onset. The N2 is significantly greater in amplitude for high-conflict cues. (*B*) Scalp topography at the peak of the N2 response showing the midline central scalp maximum of the response (blue colors). The nose is at the top of the figure of the scalp, and left is on the left of the image. The small red circles are the locations of the electrodes. (See figure 16.4.)

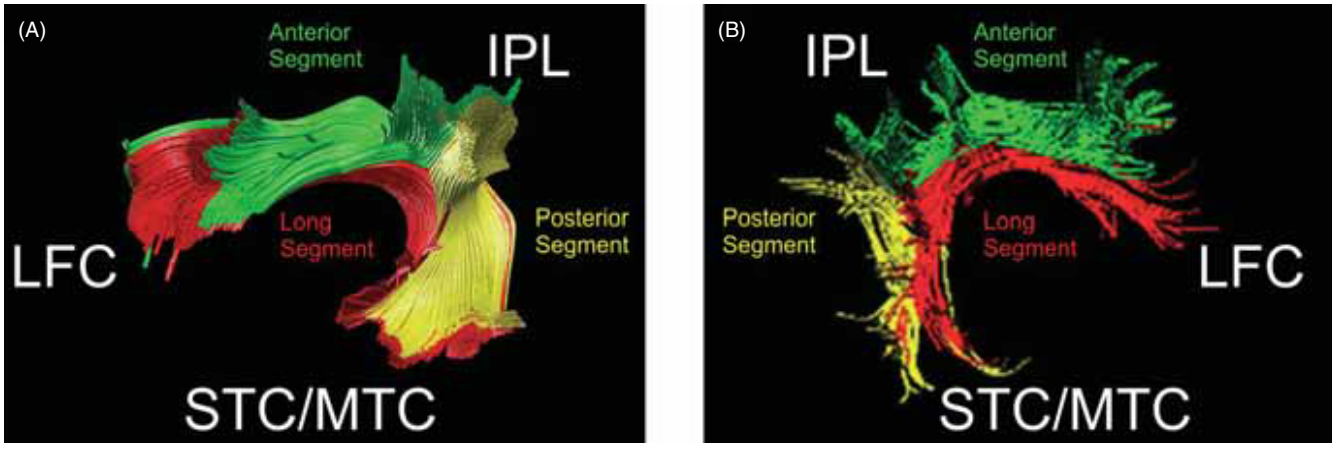


PLATE 19 Averaged tractography reconstruction for fiber connections between the superior/middle temporal, inferior parietal, and lateral frontal cortices by using a two-region-of-interest approach in (*A*) the human left hemisphere (Catani, Jones, & ffytche, 2005) and (*B*) the human right hemisphere (Gharabaghi et al., 2009). A long connection was observed linking superior/middle temporal and lateral frontal cortices (shown in red). Two shorter pathways also were found. The posterior segment running from the superior/middle temporal to the inferior parietal cortex is shown in yellow. The anterior segment running from the inferior parietal to the lateral frontal cortex is shown in green. IPL, inferior parietal lobule; LFC, lateral frontal cortex; STC, superior temporal cortex; MTC, middle temporal cortex. (With modifications from Catani et al., 2005, and from Gharabaghi et al., 2009.) (See figure 17.1.)

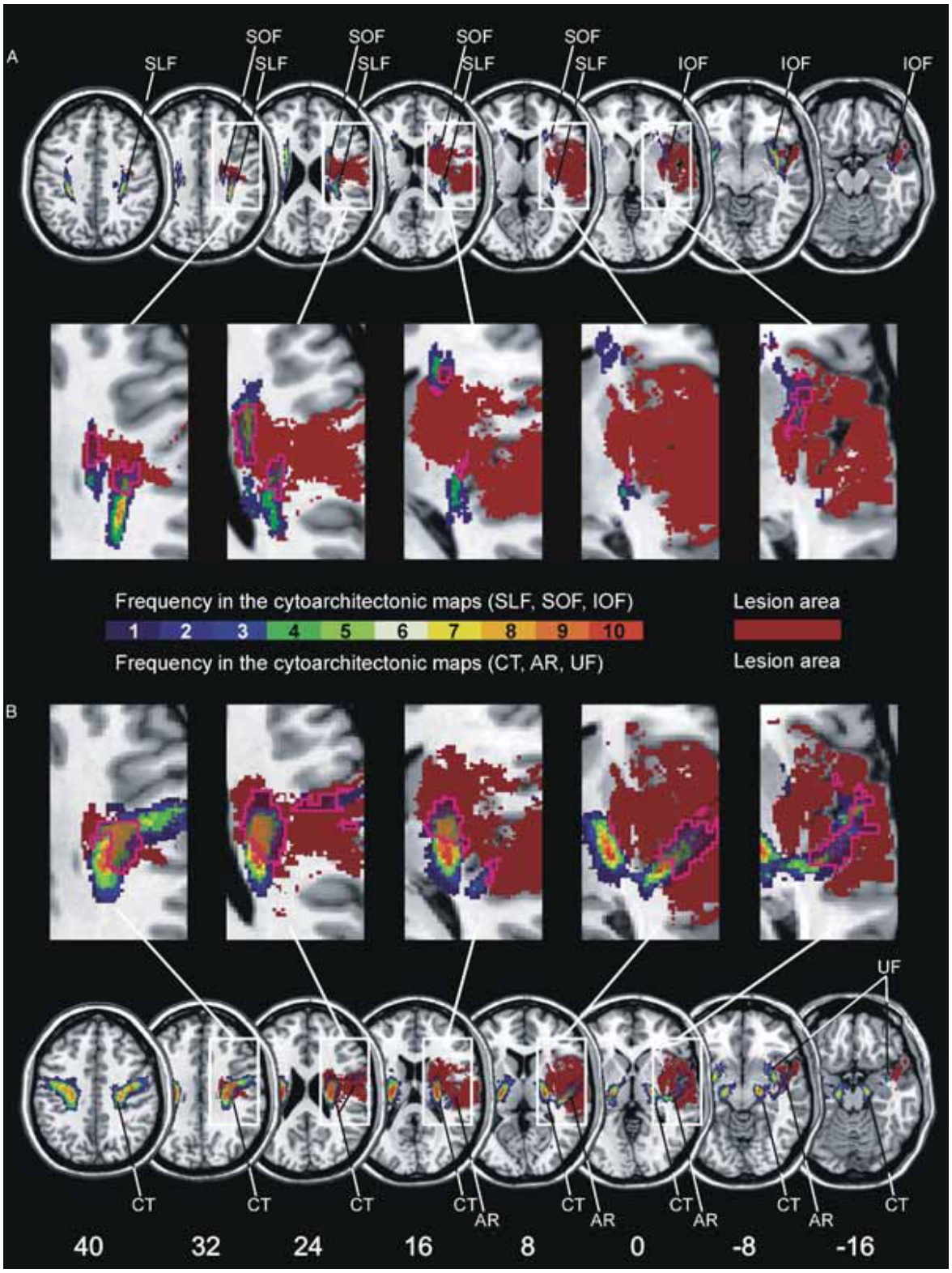


Plate 20 Overlap of the statistical VLBM lesion map (the brain territory significantly more affected in 78 patients with spatial neglect than in 62 stroke patients without this disorder) with the probabilistic, cytoarchitectonic maps of the white matter association fiber tracts from the Jülich atlas. The statistical lesion map is illustrated in homogeneous brown color. The color coding of the Jülich atlas from 1 (dark blue, observed in 1 postmortem brain) to 10 (red, overlap in all ten postmortem brains) represents the absolute frequency for which in each voxel of the brain a respective fiber tract was present (e.g., yellow color indicates that the fiber tract was present in that voxel in seven out of ten postmortem brains). The pink contour demarks the area of the fiber tracts affected by the statistical lesion map. (*A*) Overlap illustrated for perisylvian fiber tracts SFL, superior longitudinal fasciculus; IOF, inferior occipitofrontal fasciculus; and SOF, superior occipitofrontal fasciculus. (*B*) Overlap illustrated for fiber tracts CT, corticospinal tract; AR, acoustic radiation; and UF, uncinate fascicle. (From Karnath et al., 2009.) (See figure 17.3.)

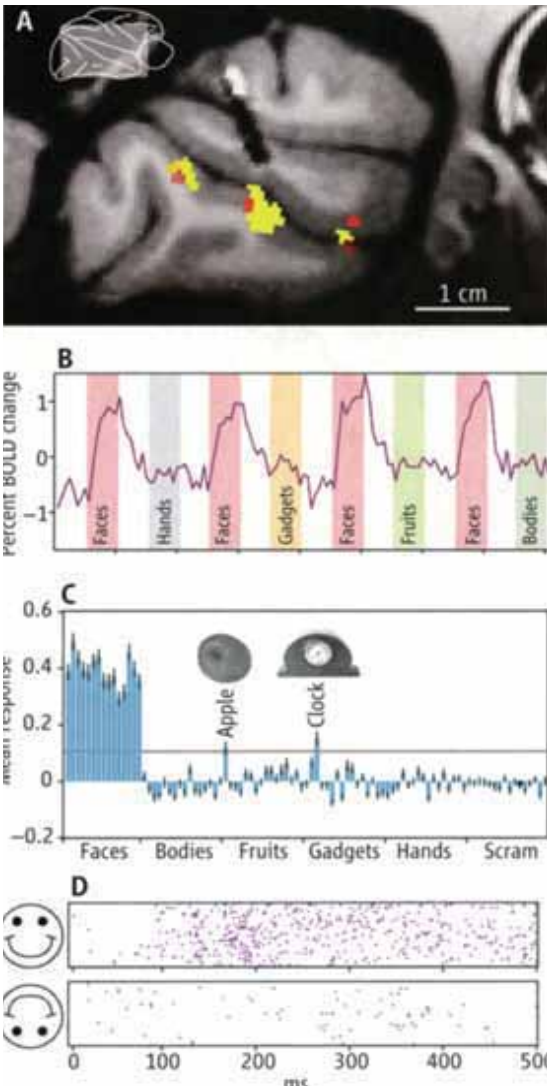


Plate 21 (*A*) Three patches of face-selective fMRI activation (yellow regions) in the macaque temporal lobe. (*B*) Time course from the face patches. Blood flows to these regions increase only when the monkey views faces. (*C*) Average response across 182 cells from the middle face patch of one monkey to 96 different images. The first 16 images are faces. (*D*) Responses of a face cell to repeated presentations of an upright and an inverted cartoon face. Each dot represents an action potential. (From Tsao, 2006.) (See figure 21.3.)

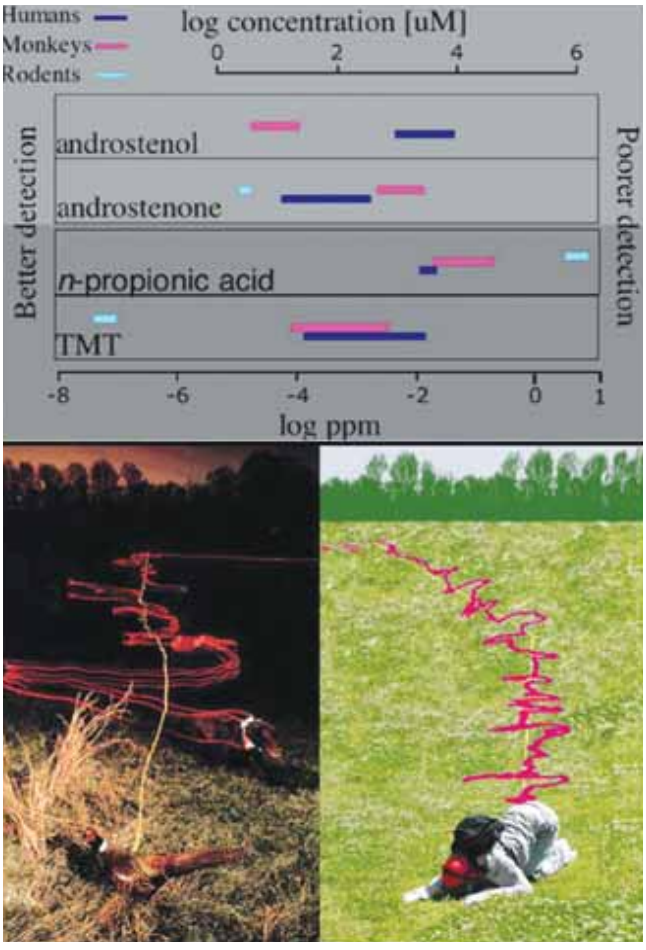


Plate 22 Olfaction in humans and other animals. *Top panel*: Detection thresholds across species. The data are amassed from studies by Laska and colleagues (1999) and are for detection of the fox odor TMT, n-propionic acid, and the two steroidal compounds androstenol and androstenone. The two shades of gray distinguish the units in which results were reported, dark gray in log concentration of the vapor, and light gray in log concentration of the odorant liquid. The extent of the lines reflects the reported variance across studies. The important point illustrated is that each species excels at detecting particular odorants. For example, humans outperform rats and monkeys at detecting n-propionic acid. That said, one must keep in mind the limitation of comparing across studies that used different methods of delivery and statistical criteria. *Bottom panel*: Human subject's path following a scent trail, as compared to a dog's path. *Left*: Path of a dog following the scent trail of a pheasant dragged through a field (scent trail in yellow, dog's path in red) (Gibbons, 1986). *Right*: Path of a human following a scent trail of chocolate essential oil through a field (scent trail in yellow, human's path in red). (The background trees were pasted in for esthetics and are part of the data.) (From Porter & Sobel, 2005.) (See figure 22.2.)

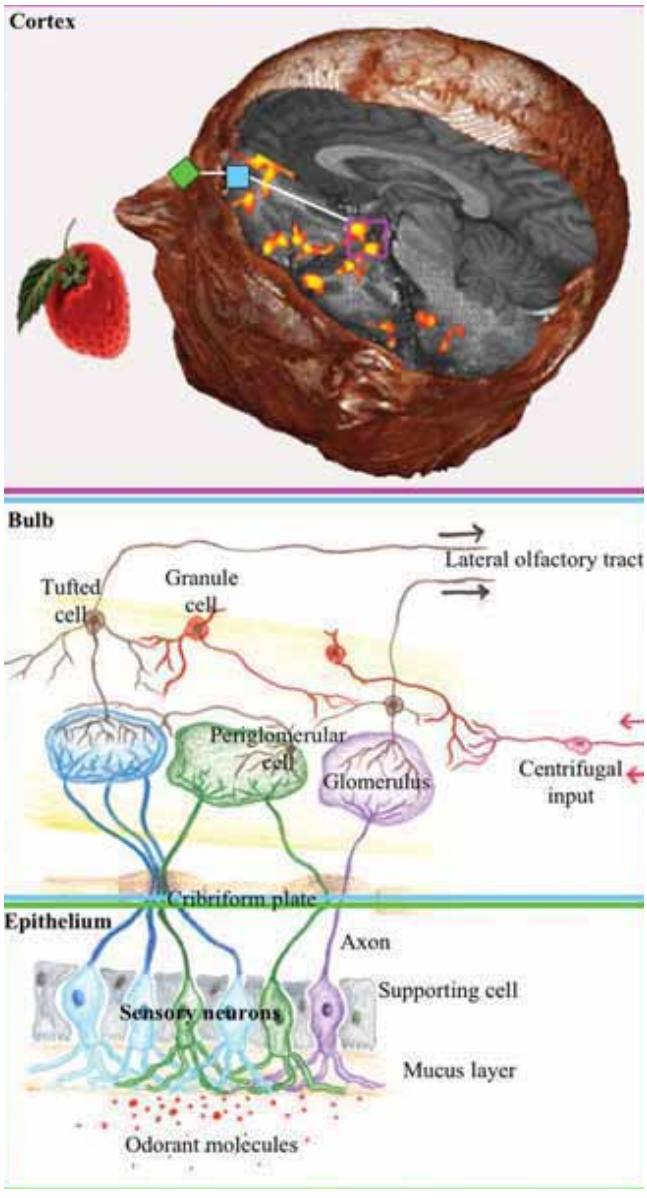


Plate 23 Structure of the human olfactory system. The human olfactory system can be segregated into three primary compartments: (bottom) epithelium, (middle) bulb, and (top) cortex. Olfactory epithelium: Each olfactory sensory neuron expresses on olfactory receptor gene. Like receptor project to one or a small number of glomeruli. Organization of the olfactory bulb: Glomeruli receive input from olfactory sensory neurons and cortical olfactory regions. Mitral and tufted cell dendrites contact receptor axons within glomeruli. The axons of the mitral and tufted cells project widely to higher brain structures. Lateral processing in the olfactory bulb occurs across two types of interneurons: periglomerular cells and granule cells. A sagittal view of the human head. The olfactory epithelium is in green, bulb in blue, and primary cortex in pink. (Drawing courtesy of Christina Zelano.) (See figure 22.3.)

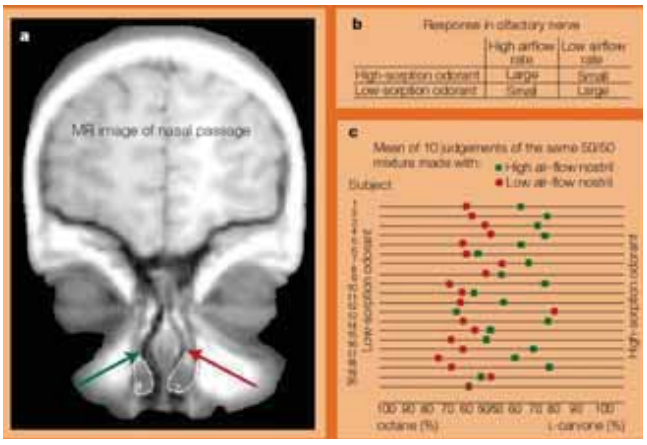


Plate 24 Different nostrils convey different olfactory information to the brain. (*A*), Magnetic resonance image of the nasal passages. The swollen (*) and relaxed (#) turbinates, outlined in white, result in an occluded right nostril (red arrow) and a clearer left nostril (green arrow). (*B*), The size of the response in the olfactory nerve (large or small) as a function of the interaction between airflow rate and odorant sorption (Mozell & Jagodowicz, 1973). (*C*), On each of 10 trials, subjects were asked to smell an identical mixture of 50% octane and 50% L-carvone using either the left or right nostril. They were then given each individual odorant component to smell separately and judged the composition of the mixture by marking the line. Using the high-flow-rate nostril (green), the average judgment was that the mixture consisted of 55% L-carvone and 45% octane. Using the low-flow-rate nostril (red), the judgment was that it consisted of 61% octane and 39% L-carvone ($t(19)43.74$, $p = 0.001$). (From Sobel et al., 1999.) (See figure 22.4.)

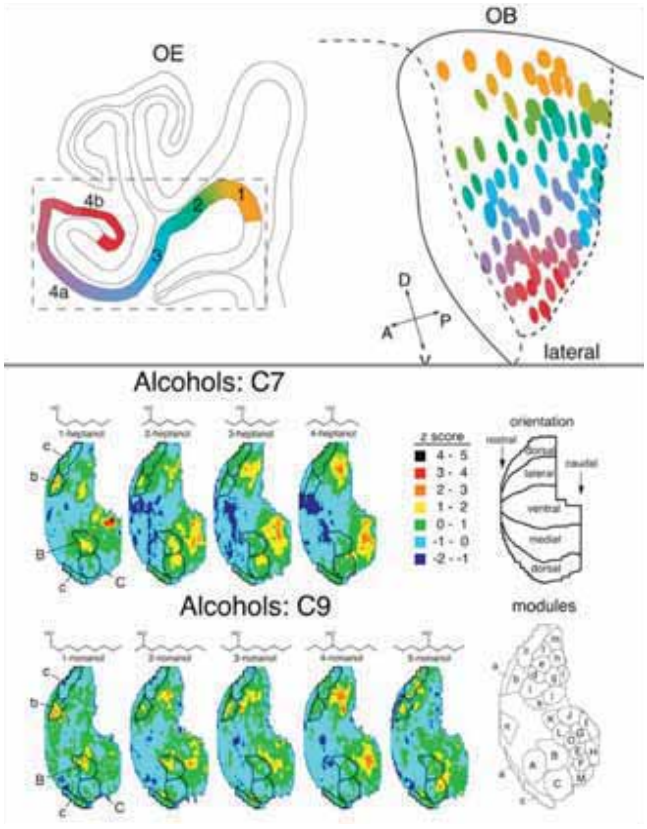


PLATE 25 Spatial mapping from epithelium to bulb. *Top panel*: Color-coded zonal distribution of receptor types in the olfactory epithelijm and their projection pattern to glomeruli in the olfactory bulb (Miyamichi et al., 2005). *Bottom panel*: Patterns of 2DG activation on the surface of the rat olfactory bulb as a reflection of odorant identity. For detailed maps, see http://leonserver.bio.uci.edu/. (See figure 22.7.)

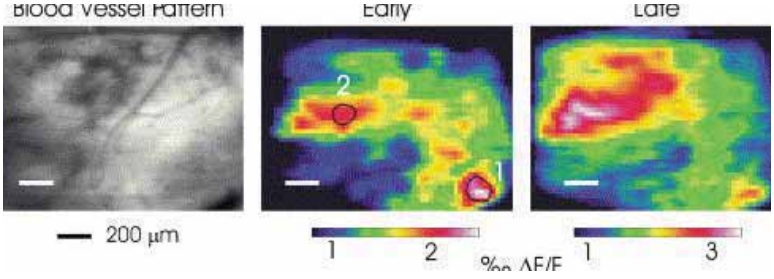


PLATE 26 Temporal development of odor-induced activity. Data from Spors and Grinvald (2002) showing the temporal development of the bulbar response to the odorant ethylbutyrate. The early response is data obtained 150–300 ms following stimulus onset, and the late response is data obtained 300–500 ms following stimulation. The spatial pattern of response is clearly modified over time. (See figure 22.8.)

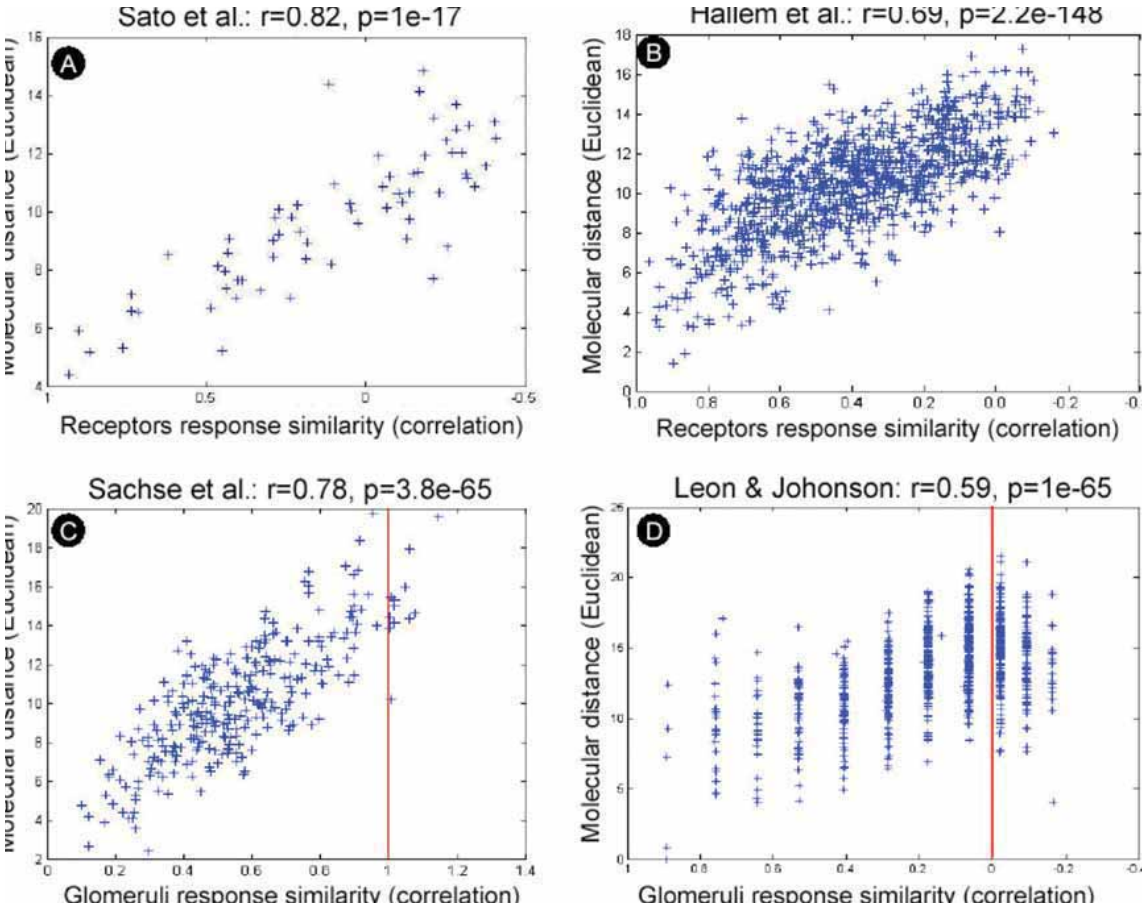


PLATE 27 Correlation plots for unrelated data sets. The graphs demonstrate the ability of the multidimensional olfactory metric to predict neural activity in the olfactory system. Graphs *A* and *B* are from data set reporting RN responses (the similarity was the measured Pearson correlation and thus can range from 1 to −1). Graphs *C* and *D* are from data sets reporting GLO responses. In the *C* and *D* data sets, the *r* value was positive as long as the response pattern similarity was between 0 to 1, and it was either negative or low when the response pattern similarity was negative (the right part of the red line). (*A*) Hallem et al. data set. (*B*) Sato et al. data set. (*C*) Sachse et al. data set (*D*) Leon and Johnson data set. (See figure 22.14.)

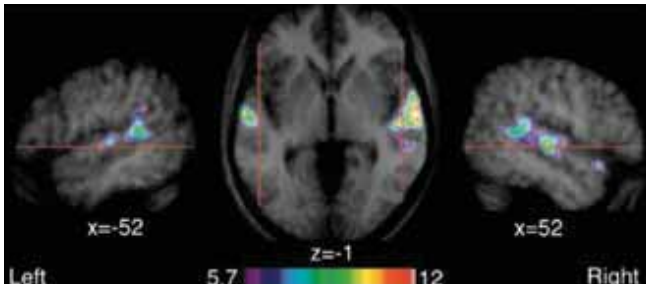


PLATE 28 Activation during passive listening to voices (Belin et al., 2000). The numbers refer to planes defined in millimeters in Talairach space (Talairach & Tournoux, 1988). (See figure 25.5.)

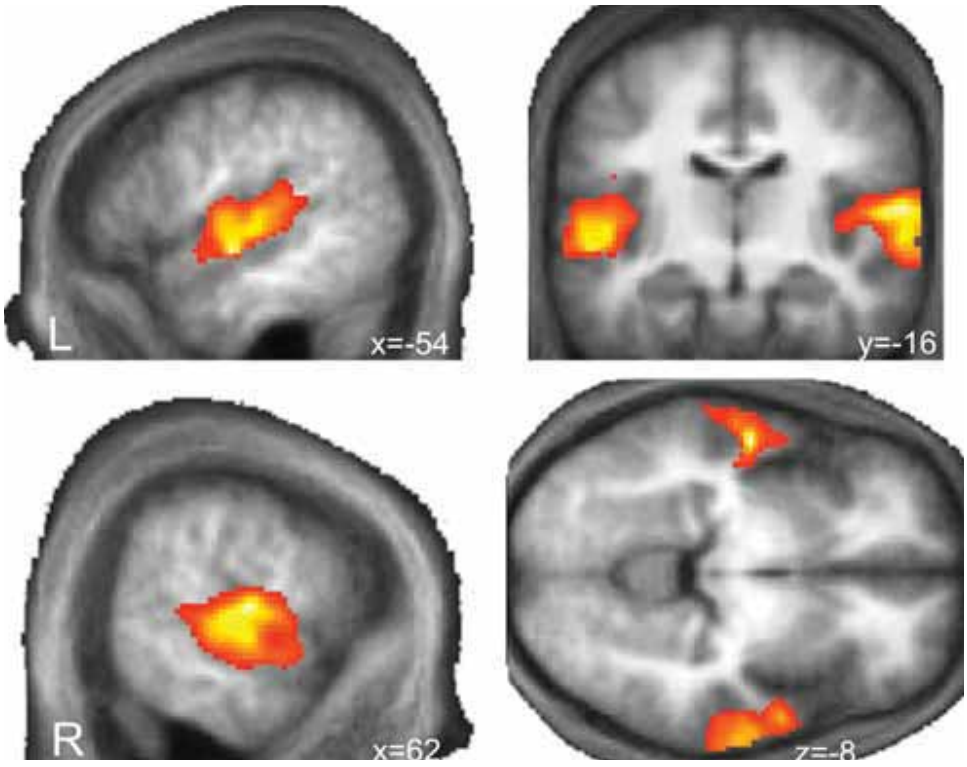


PLATE 29 Activation due to passive listening to changing resonator scale and sound class in the three types of harmonic sounds shown in figure 25.2 (von Kriegstein et al., 2007). (See figure 25.6.)

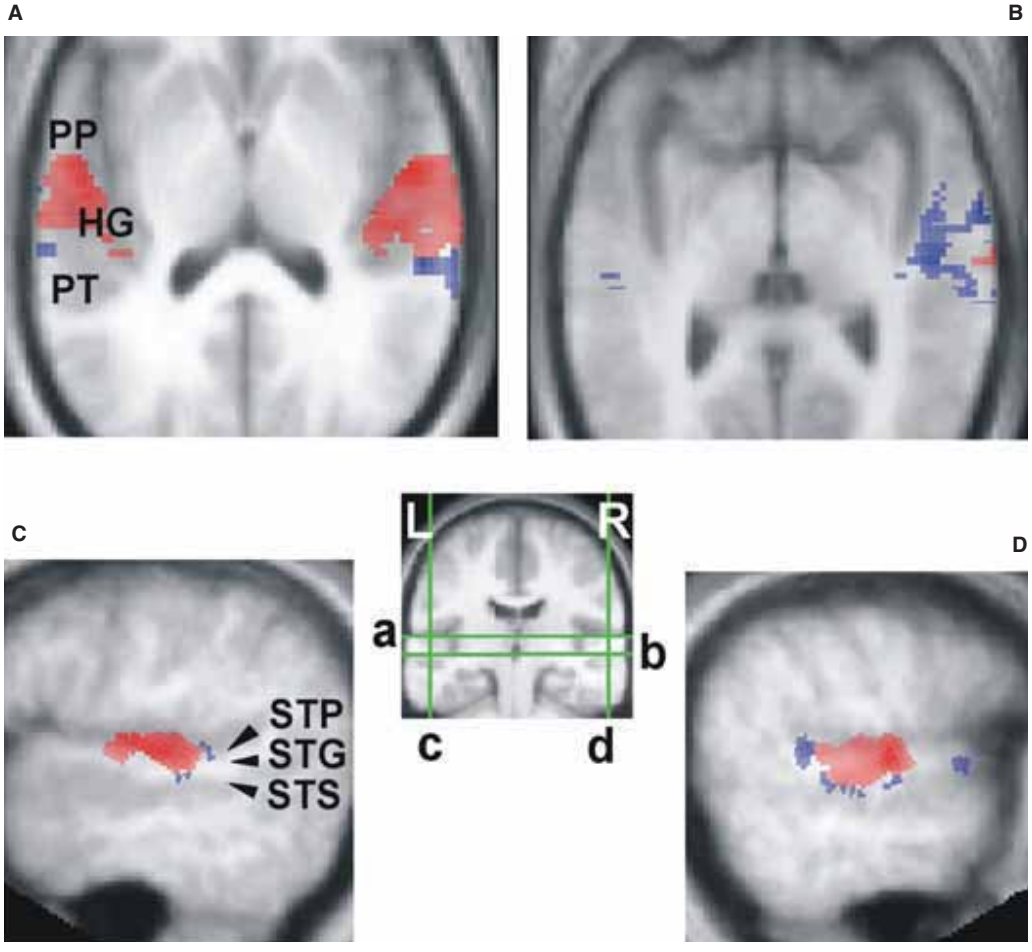


PLATE 30 Activation due to passive listening to changing spectral envelope (Warren et al., 2005). HG, Heschl's gyrus; PT, planum temporale; PP, planum polare; STS, superior temporal sulcus. (See figure 25.7.)

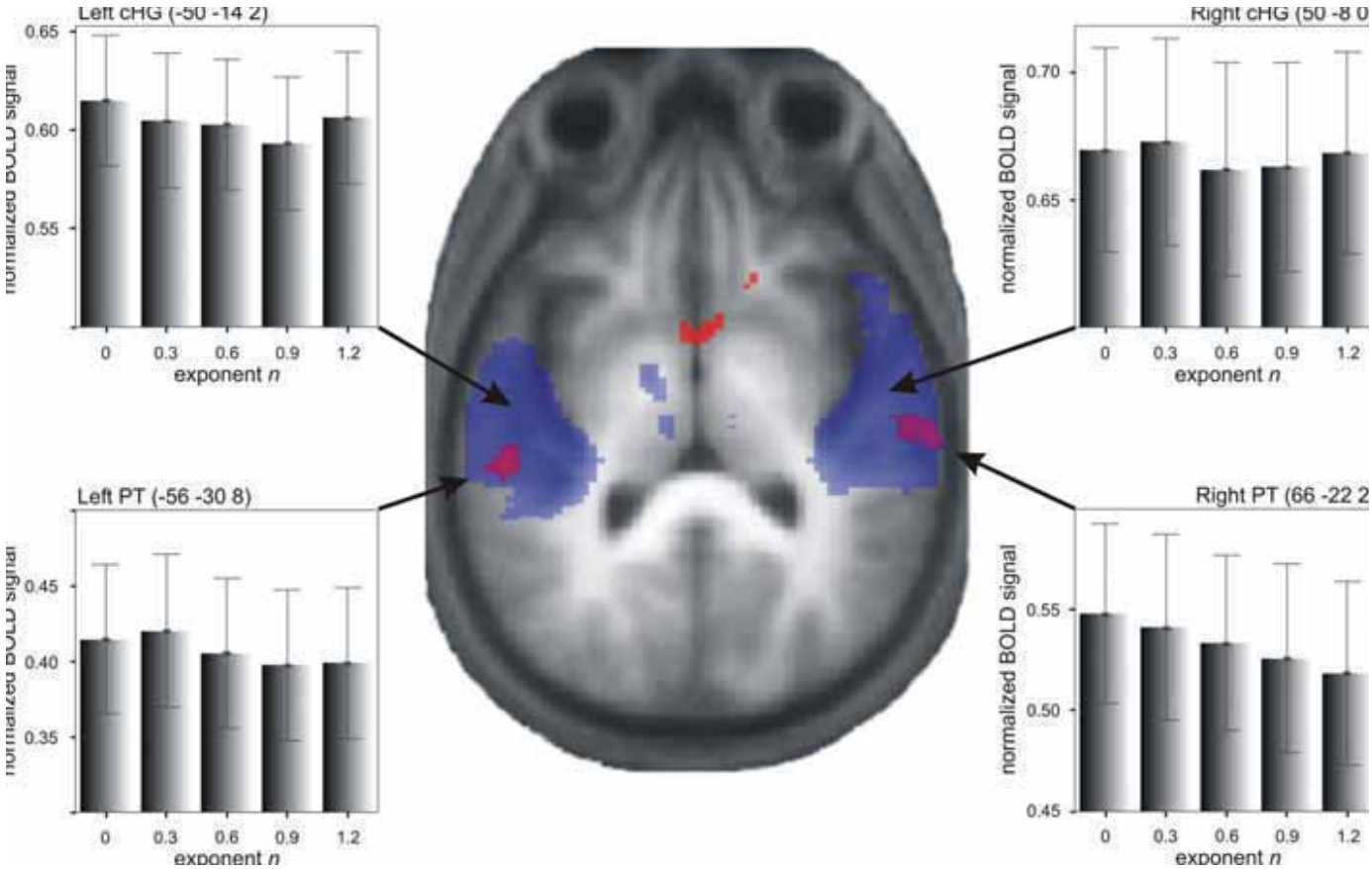


PLATE 31 Activation as a function of information content of pitch sequences during the encoding of pitch sequences. A significant effect of pitch-sequence information content (which decreases as *n* increases) is shown in the planum temporale (PT) but not in Heschl's gyrus (HG). Numbers in parentheses are Talairach coordinates in millimeters where the BOLD values were measured. (See figure 25.8.)

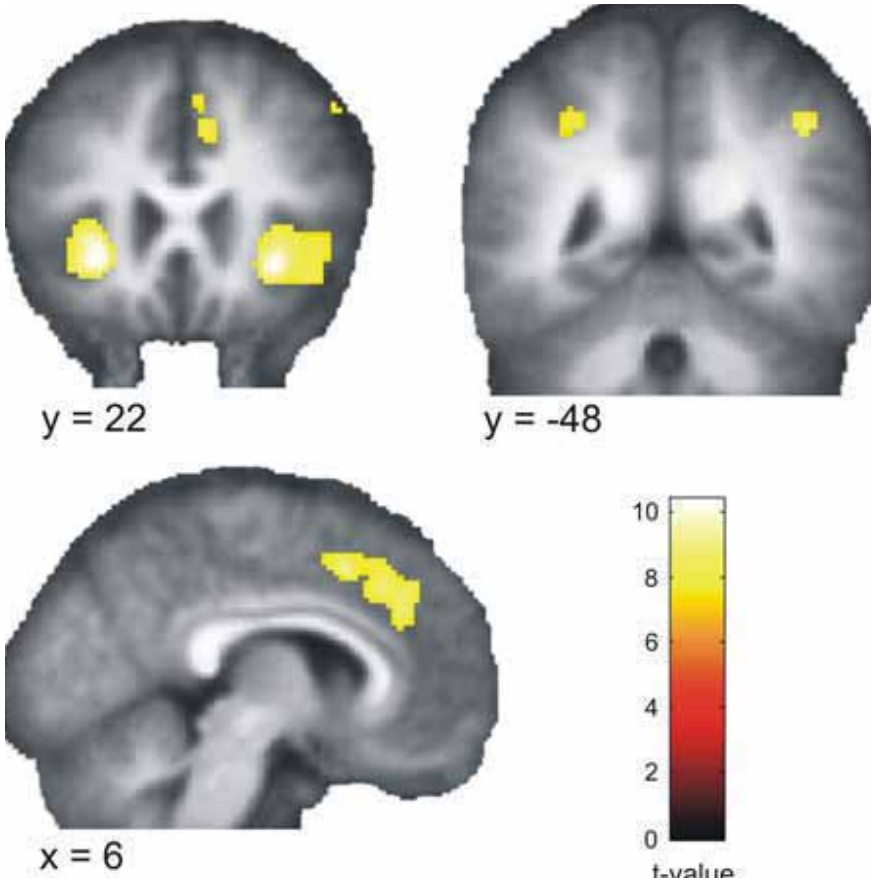


PLATE 32 Contrast to demonstrate activation associated with pitch-sequence retrieval in an active listening task. (See figure 25.9.)

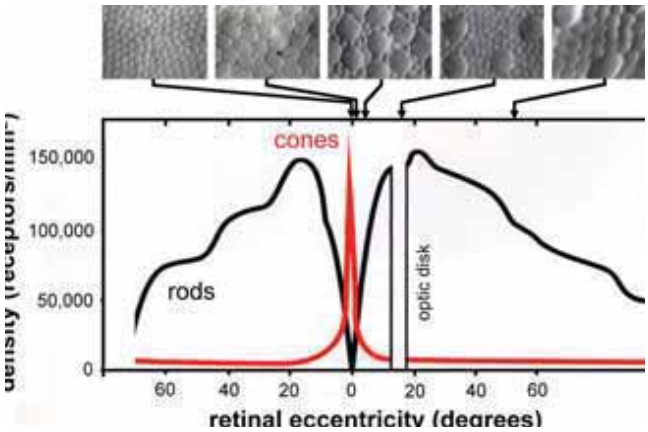


PLATE 33 Nonuniform distribution of rods and cones in the human retina. Plot of photoreceptor density as a function of retinal eccentricity. The top panels show ex vivo images of the photoreceptor mosaic from Curcio, Sloan, Kalina, and Hendrickson (1990). The leftmost image is from the all-cone fovea; the remaining panels contain both rods (smaller cells) and cones (larger cells). While rod density increases dramatically in the peripheral retina, rod diameter remains relatively constant (about 2 μm). Conversely, the cone photoreceptors increase from about 2 μm in diameter at the fovea to about 8 μm at about 10 degrees eccentricity, after which point they remain relatively constant (Samy & Hirsch, 1989). (Modified from Webvision (http://webvision.med.utah.edu), with permission.) (See figure 26.1.)

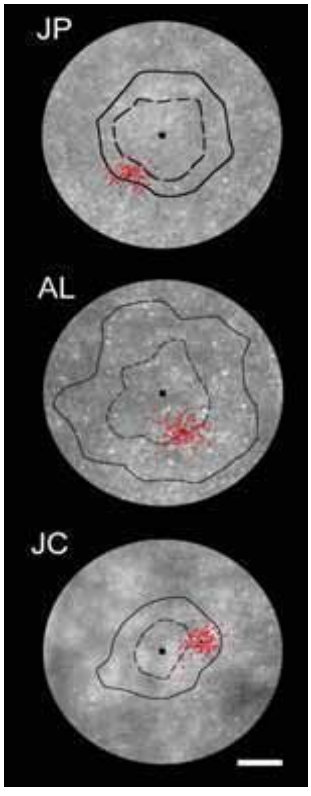


PLATE 34 The area of highest cone density is not always used for fixation. Shown are retinal montages of the foveal cone mosaic for three subjects. The black square represents the foveal center of each subject, as defined by the location of peak cone density. The dashed black line is the isodensity contour line representing a 5% increase in cone spacing, and the solid black line is the isodensity contour line representing a 15% increase in cone spacing. Red dots are individual fixation locations. Scale bar is 50 μm. (Reproduced from Putnam et al., 2005, with permission.) (See figure 26.2.)

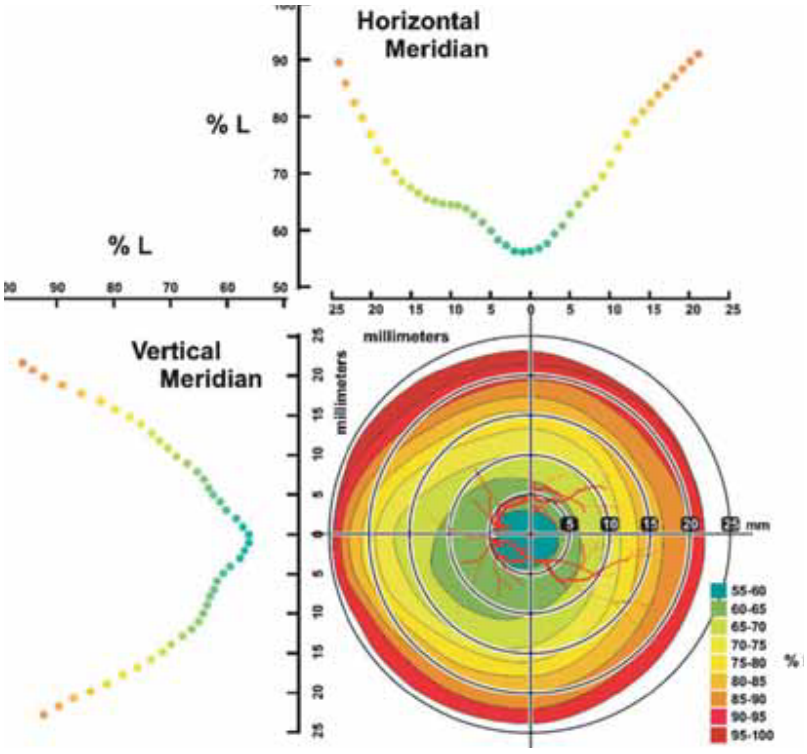


PLATE 35 The L-to-M cone ratio is not constant across the retina. The lower right panel shows a topographical map of the percent of L-opsin mRNA in a human donor retina. The proportion of L-opsin to M-opsin mRNA is directly related to the relative numbers of L and M cones at a locus, assuming that L and M cones produce the same absolute amounts of mRNA. Horizontal and vertical meridian slices show the dramatic increase in percent of L as a function of eccentricity. (Reproduced from Neitz et al., 2006, with permission.) (See figure 26.4.)

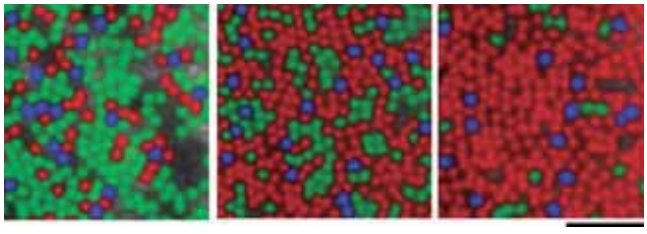


PLATE 36 Intersubject variation in L-to-M cone ratio. False color images showing the arrangement of L (red), M (green), and S (blue) cones in the retinas for three human subjects. The identity of each cone as L, M, or S was inferred from retinal densitometry measurements obtained with an adaptive optics fundus camera. The proportion of S cones does not vary significantly between subjects; however, the L-to-M cone ratio can vary by a factor of 40 across individuals with normal color vision. Scale bar is 5 arcmin. (See figure 26.5.)

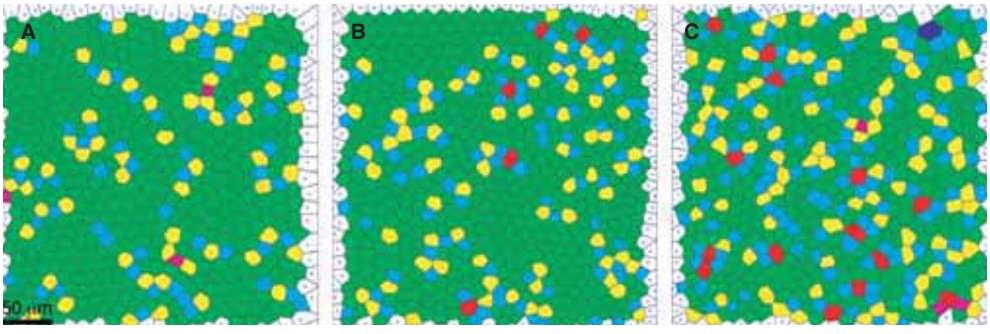


PLATE 37 Regularity of the human cone mosaic. Voronoi domain associated with each cone photoreceptor in a patch of retina from (*A*) a normal trichromat, (*B*) a 34-year-old female with a mild tritan defect, and (*C*) a 57-year-old male with a severe tritan defect. The color code indicates the number of sides on each Voronoi polygon (magenta = 4, cyan = 5, green = 6, yellow = 7, red = 8, purple = 9). Large regions of six-sided polygons indicate a regular triangular lattice, whereas other colors mark points of disruptions in the hexagonal packing of the foveal mosaic. Despite the fact that the father and the daughter carried the same heterozygous mutation in their S-opsin genes (predicting a tritan phenotype), the regularity of the father's mosaic was significantly disrupted, while the daughter's was indistinguishable from normal. Scale bar is 50 µm. (Reproduced from Baraas et al., 2007, with permission.) (See figure 26.6.)

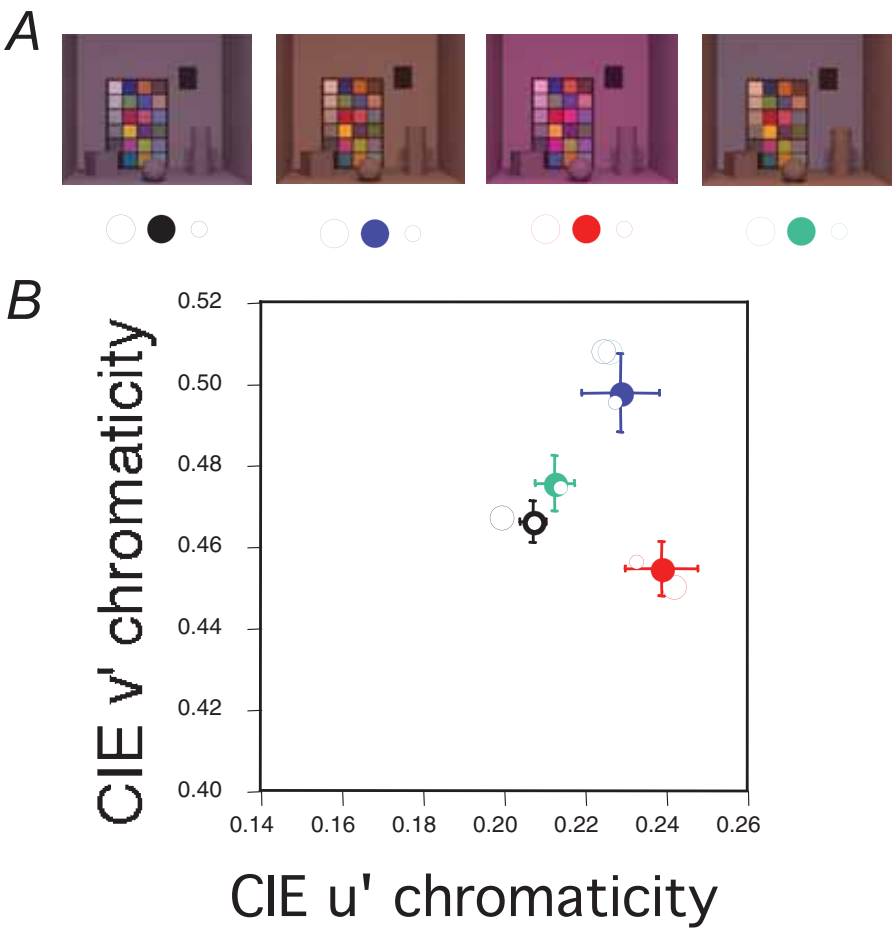


PLATE 38 Color constancy performance. (*A*) Images of 4 of 17 simulated scenes used to compare human performance and a model derived from a Bayesian illuminant estimation algorithm. Each scene has the same spatial structure. In the first three images, from left to right, the illuminant varies. In the rightmost image, the illuminant is the same as that in the second image, but the background surface has been changed so that the light reflected from it matches that reflected from the background surface in the leftmost image. (Reproduced from Brainard et al., 2006, figure 1.) (*B*) Illuminants, achromatic loci, and model predictions plotted in the CIE $u'v'$ chromaticity diagram. This is a standard color representation that preserves information about the relative responses of the L, M, and S cones but not about intensity. Large open circles show the scene illuminants, with the color key as indicated beneath the images in panel *A*. Large solid circles show the achromatic loci measured by observers who adjusted a test patch at the location indicated by the black rectangle in each image. The small open circles show the model's predictions of the achromatic loci. (Reproduced from figure 7 of Brainard et al., 2006.) (See figure 27.4.)

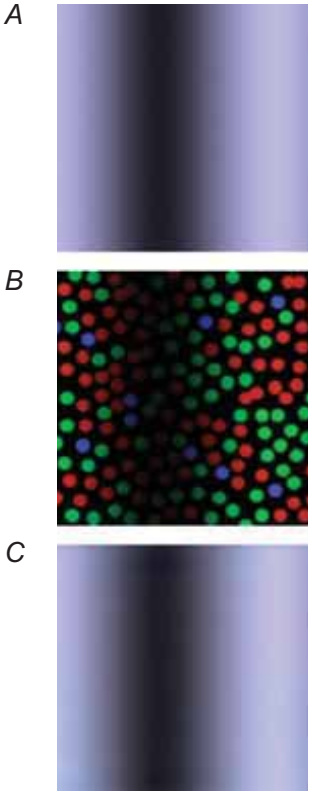


PLATE 39 Trichromatic reconstruction. (*A*) Small patch of sinusoidal isochromatic grating. (*B*) The intensity of each colored spot represents the isomerization rate of a cone. The class of each cone is indicated by whether it is plotted in red (L), green (M), or blue (S). (*C*) The isochromatic grating as reconstructed by the Bayesian algorithm. The grating shown corresponds to a spatial frequency of 6 cycles per degree presented at about 1 degree of eccentricity for a human observer; the mosaic is of observer AP of Hofer and colleagues (2005). Brainard and colleagues (2008) provide additional reconstruction examples that show similarly veridical performance for low-spatial-frequency isoluminant gratings and for an additional mosaic. (See figure 27.5.)

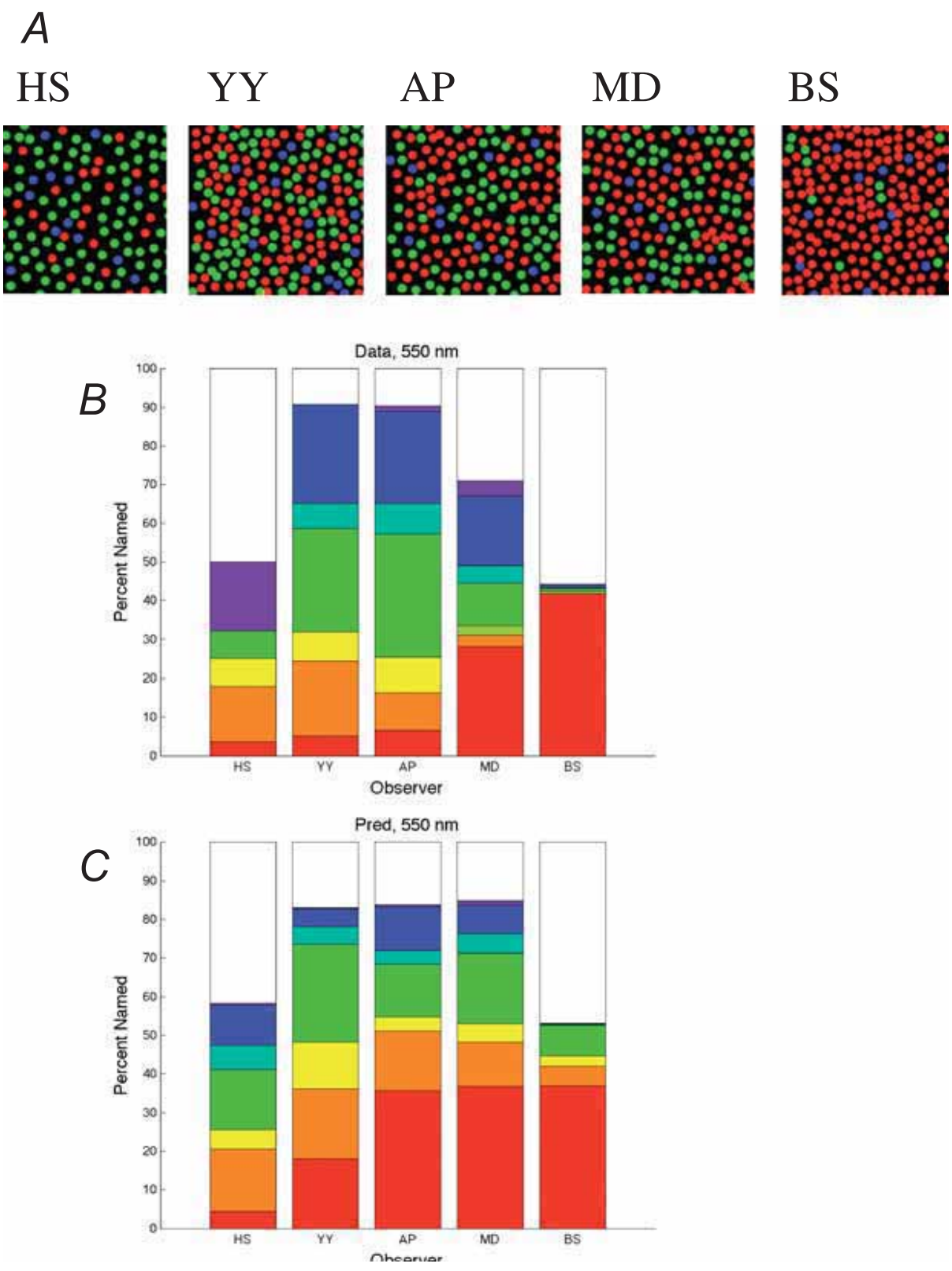


PLATE 40 Small spot experiment. (*A*) Schematic of five individual observer cone mosaics. Red, green, and blue circles show locations of L, M, and S cones. Mosaics represent approximately 12 by 12 feet of visual angle at 1 degree of eccentricity. (Reproduced from Brainard et al., 2008, figure 2 (top panel).) (*B*) Data from Hofer and colleagues (2005) for 550-nm spots. Observers named each small spot that they saw and judged namable. The available names were red, orange, yellow, yellow-green, green, blue-green, purple, and white. For each observer, the histogram shows the proportion of each color named used, with the color code corresponding to the name. Note that the white region of the bars represents the proportion of white responses. Not all observers used all available names. (Reproduced from figure 11 of Brainard et al., 2008.) (*C*) Predictions from the Bayesian model, obtained as described in the text, for the experimental conditions corresponding to the data in *B*. (Reproduced from figure 11 of Brainard et al., 2008), which also shows data and predictions for 500-nm and 600-nm spots.) (See figure 27.6.)

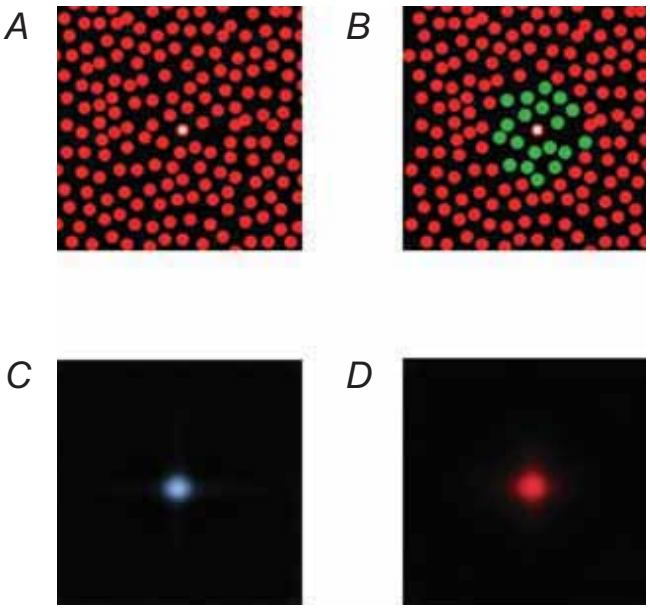


PLATE 41 Small spot intuitions. (*A*) A mosaic consisting only of L cones. The white spot in the center indicates a single cone whose stimulation was simulated. (*B*) This mosaic is identical to the one shown in panel *A*, with the exception that the cones surrounding the central L cone have been changed to M cones. (*C*) Model output when the central cone in panel *A* is stimulated. The result is a bluish-white spot. As described by Brainard and colleagues (2008), a windowing procedure was applied to model output here and in panel *D* to reduce visible ringing in the reconstruction. (*D*) Model output when the central cone in panel *B* is stimulated. The result is a reddish spot. (Reproduced from figure 7 of Brainard et al., 2008.) (See figure 27.7.)

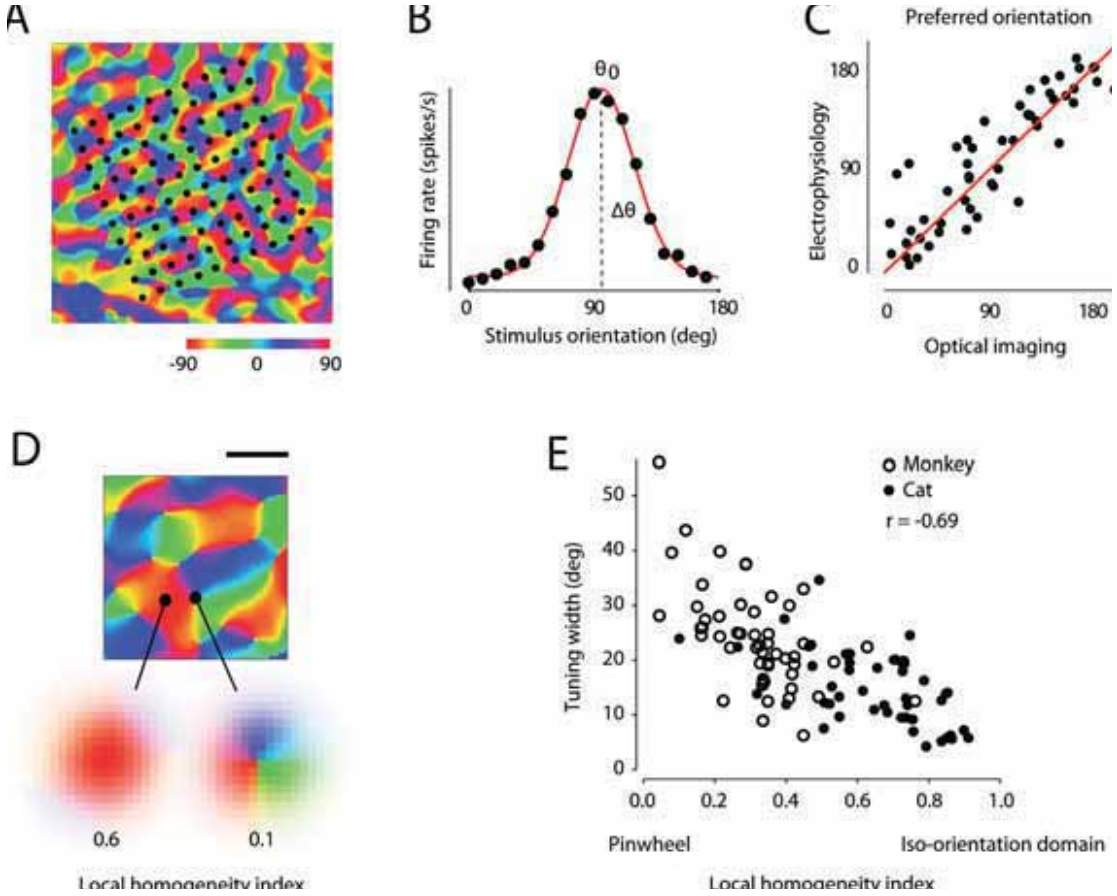


PLATE 42 Orientation tuning bandwidth and local map structure. (*A*) Example of an orientation preference map in macaque visual cortex along with the recovered location of the microelectrode array. (*B*) Reverse correlation in the orientation domain (Ringach et al., 1997) was used to measure the tuning curves at each electrode site simultaneously. The example here shows the average spike rate triggered to the presentation of each orientation in a rapid stimulus sequence, yielding a preferred orientation, θ_0, and tuning width, $\Delta\theta$. (*C*) The estimated location of the array (solid dots in panel *A*) was estimated by finding the optimal translation/rotation parameters for which the preferred orientations as measured via reverse correlation matched those measured optically. The scatterplot illustrates the optimal correlation in one instance. (*D*) A local homogeneity index was defined to capture the diversity of orientation preferences around each cortical point. The example illustrates two locations with a low homogeneity index of 0.1 attained near a pinwheel and a location with a high index of 0.6 in an iso-orientation domain. (*E*) Spatial distribution of the local homogeneity index for the same patch of cortex as the one shown in panel *D*. (*F*) Isolation of single units. Only units that could be very well isolated, as is typical of the principal component analysis here, were used in our analyses of tuning bandwidth and local map structure. (See figure 28.2.)

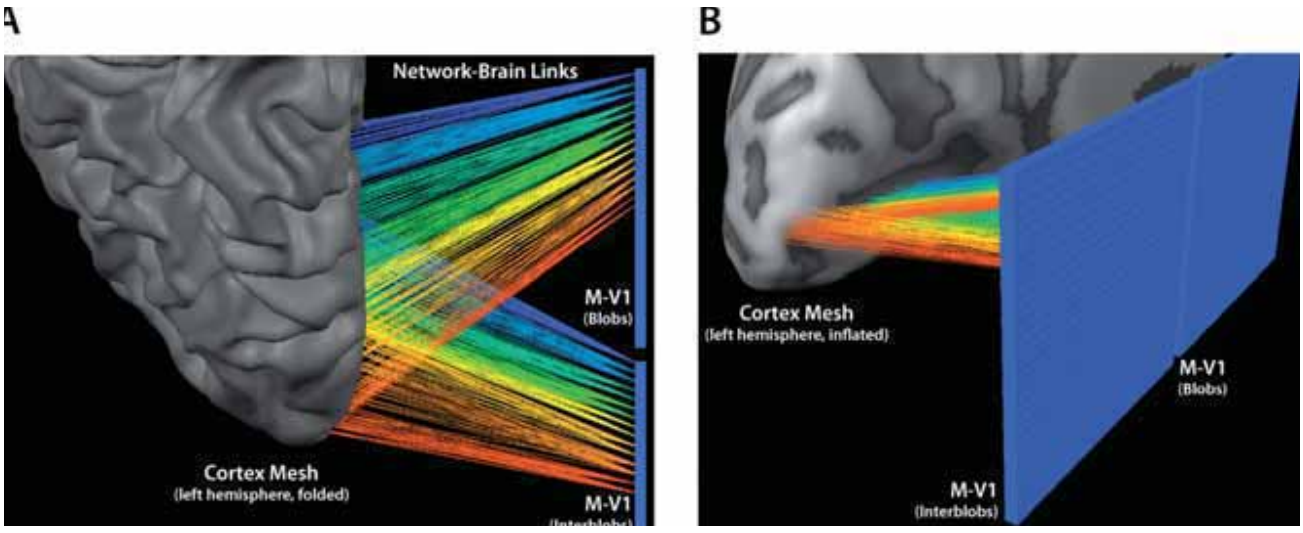
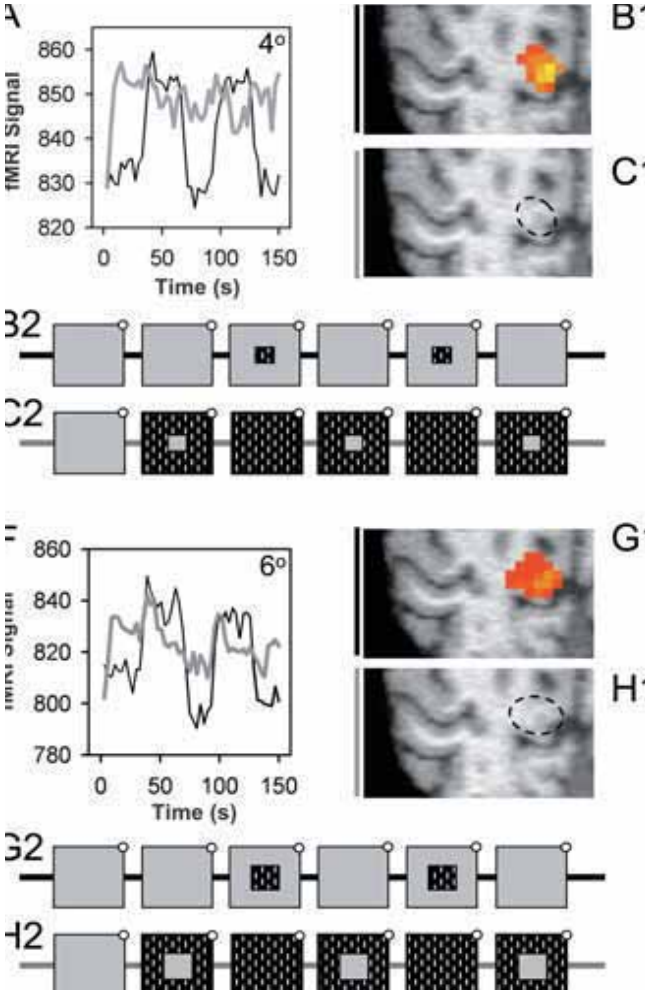


PLATE 43 Selected network-brain (NB) links connecting cortical column units from one layer of the boundary processing system (interblobs) and one layer of the surface processing system (blobs) of M-V1 to a left hemisphere cortex mesh of a subject. Colored lines indicate NB links from foveal (red) to increasingly peripheral (orange, yellow, green, blue) locations within the model layers and corresponding regions within the calcarine sulcus of the subject's cortex as identified by a fMRI retinotopic mapping experiment. (*A*) Top view depicting the cortex mesh in a folded state. (*B*) Lateral view depicting the cortex mesh in an inflated view. Note that NBLs originating from the same retinotopic position of the boundary-processing M-V1 layer and surface-processing M-V1 layer are connected to the same position at the cortex mesh. (See figure 30.4.)



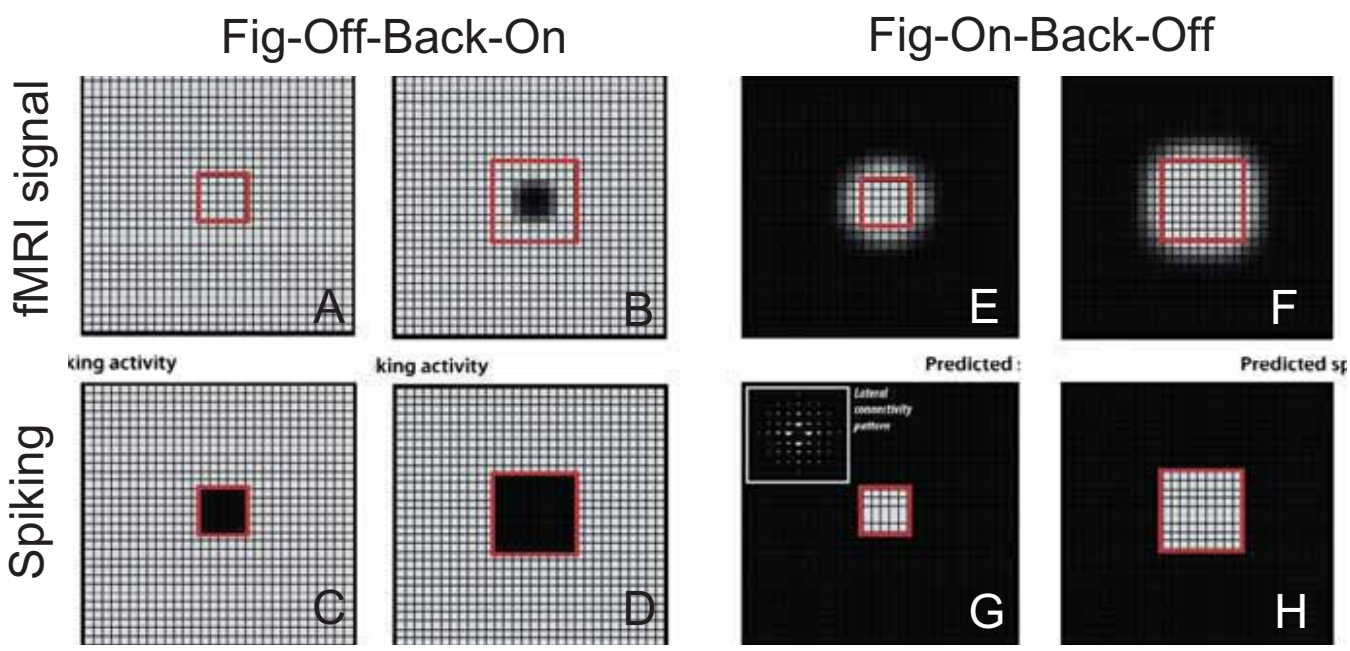


PLATE 45 Simulation of spiking activity and fMRI signals when an "empty" square is surrounded by dynamic texture (Fig-Off-Back-On) or vice versa (Fig-On-Back-Off). As indicated by red outlines, stimuli were presented with a small square (5 × 5 rectangle; *A*, *C*, *E*, *G*) or a large square (9 × 9 rectangle; *B*, *D*, *F*, *H*). Each panel shows the activity state of the same layer from the surface-processing system of M-V2. The lateral connectivity pattern of each unit within this layer is shown in G. The activity state of a processing unit is indicated by a black-to-white color range corresponding to weak-to-strong activity. The predicted fMRI data in the Fig-Off-Back-On condition show inflow from texture background into the figure representation (*A*, *B*), and only for the larger square (*B*) the interior is spared (shown in dark). Predicted spiking data (*C*, *D*) show a perfect representation of the square, without noticeable inflow from the background. The predicted fMRI data in the Fig-On-Back-Off condition show outflow of fMRI activity from the representation of the texture square into the empty background (*E*, *F*), while such outflow is unnoticeable for spiking data (*G*, *H*). (See figure 30.6.)

PLATE 44 Empirical data from a single subject illustrating limitations in fMRI resolution using 1.5 Tesla Signa Horizon Echospeed system. Similar data were obtained from a second subject, and data in both subjects were replicable across sessions. Functional scans were obtained with a gradient echo EPI sequence (BOLD images), using a 64 × 64 matrix, a FOV of 14–16 cm, coronal slices with Thk = 4 mm, TE = 40 ms, and TR = 4 s. Functional data were overlaid on high-resolution structural scans of the same person's brain. Structural scans were obtained by using 3D-SPGR, a 512 × 384 × 128 matrix, with TE = 5 ms, TR = 24 ms, and a flip angle of 45 degrees. The testing of the effects of interest resulted in Wilkinson's maps, which were converted into *z*-maps. Single voxels were considered significant when the corresponding *z*-score exceeded 3.07. The coronal slice shown was positioned 16 mm anterior from the occipital pole. Dynamic texture stimuli (see figure 30.1*A*) were equiluminant with the background (24 cd/m^2). (*A–C*) fMRI signal as a function of time (*A*) during two block designs (*B*,*C*). In the first design (*B*2), two 30-s blocks of presentation of 4 degree square at 7.2 degrees eccentricity in lower left quadrant were interleaved with three 30-s periods of gray background. In the second design (*C*2), two 30-s blocks of presentation of a textured background were alternated with three 30-s periods of presentation of a gray 4-degree square at 7.2 degrees eccentricity. An initial period of baseline measurement without stimulus was discarded from analysis. Fixation spot (where subjects performed a demanding T/L discrimination task) is indicated at the top right of each stimulus panel. When the square was defined by dynamic texture on a grey background (*B*2), significant activity (black plot in *A*) was found in a large number of voxels in the upper bank of the calcarine sulcus (B1) using a regressor corresponding to the timing of the texture square in the block design in panel *B*2. When a full texture background was shown in blocks with or without a gray square (*C*2), there was no significant activity (*C*1) for the regressor corresponding to the physical filling-in of the gray square with texture (*C*2). The gray plot in panel *A* shows activity in the region of interest (ROI) defined by the response to the textured square (ROI is shown as a dashed oval in *C*1) during data from the block design shown in *C*2. For the design in *C*2, the data suggest spread of fMRI signal from the background texture into the gray figure. (*F–H*) Same conventions as in panels *A–C*, but the square size was 6 degrees. The gray plot in panel *F* shows a small, transient response to physical filling-in of the texture (design *H*2) that did not lead to significant activity in *H*1 (based on regressor corresponding to design H2).The fMRI signal in designs *C*2 and *H*2 inside the figure representation was not due to perceptual filling-in, as the signal was present from the beginning of stimulation (gray plot in panels *A* and *F*). Based on data in block designs showing a texture square on a gray background alternated with a gray background (as in *B*2 and *G*2), we found (averaged over subjects) activated regions of 158, 230, and 267 mm^2 for square sizes of 1, 4, and 6 degrees, respectively, while based on retinotopy (Sereno et al., 1995), activated regions of approximately 3, 45, and 110 mm^2 were expected. Averaged over subjects and conditions, a Gaussian filter of 7 mm (HWHM) was required to simulate the blurring of expected signal by fMRI. (See figure 30.5.)

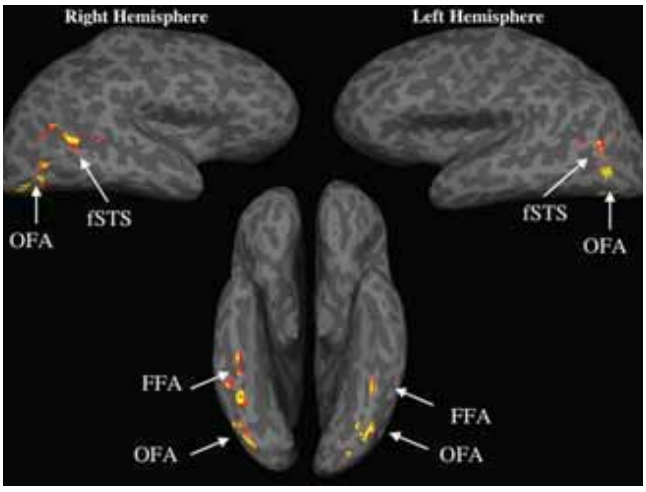


PLATE 46 Face-selective activation (faces > objects, $p < 0.0001$) on an inflated brain of one adult subject, shown from lateral and ventral views of the right and left hemispheres. Three face-selective regions are shown: the FFA in the fusiform gyrus along the ventral part of the brain, the OFA in the lateral occipital area, and the fSTS in the posterior region of the superior temporal sulcus. For studies of face identification (rather than expression, etc.), the FFA and OFA are of greatest interest. (See figure 32.1.)

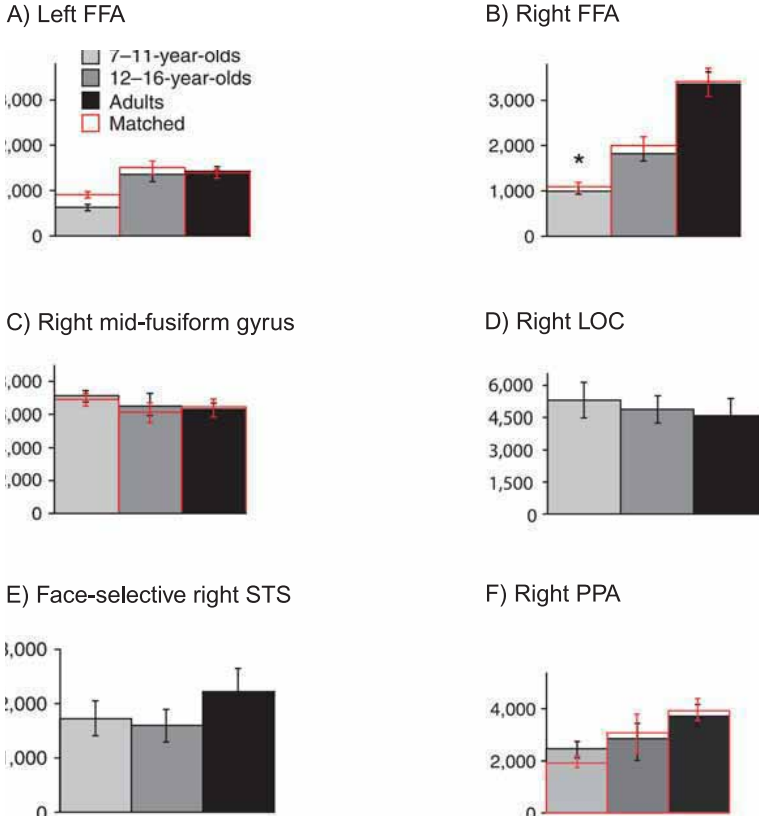


PLATE 47 Mean volume across subjects in each age group of individually defined (*A*) left and (*B*) right FFA, (*C*) anatomically defined right mid-fusiform gyrus, (*D*) functionally defined right LOC, and functionally defined (*E*) face-selective right STS and (*F*) right place-selective PPA. Red bars indicated values in subsets of subjects matched for BOLD-related confounds. (From Golarai et al., 2007.) (See figure 32.4.)

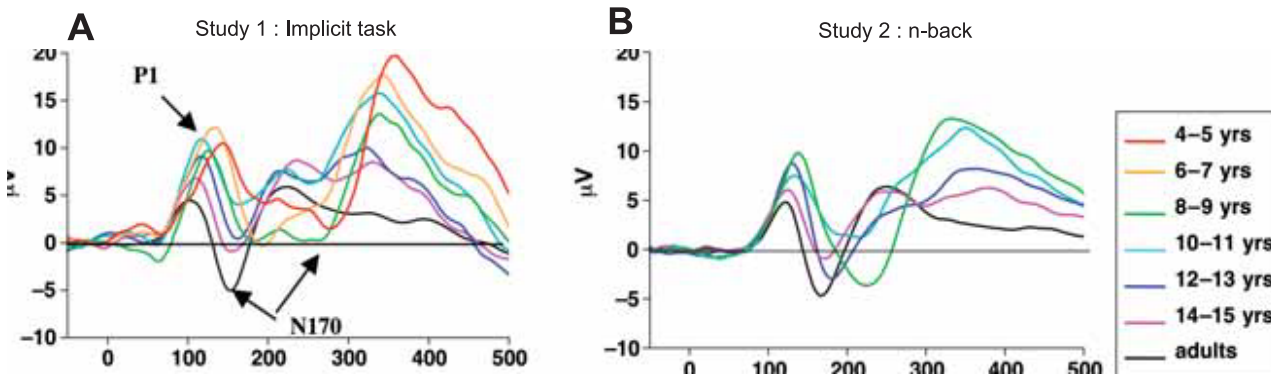


PLATE 48 ERPs from right posterior temporal scalp locations in response to face stimuli, separately for each age group. (From Taylor et al., 2004.) (See figure 32.5.)

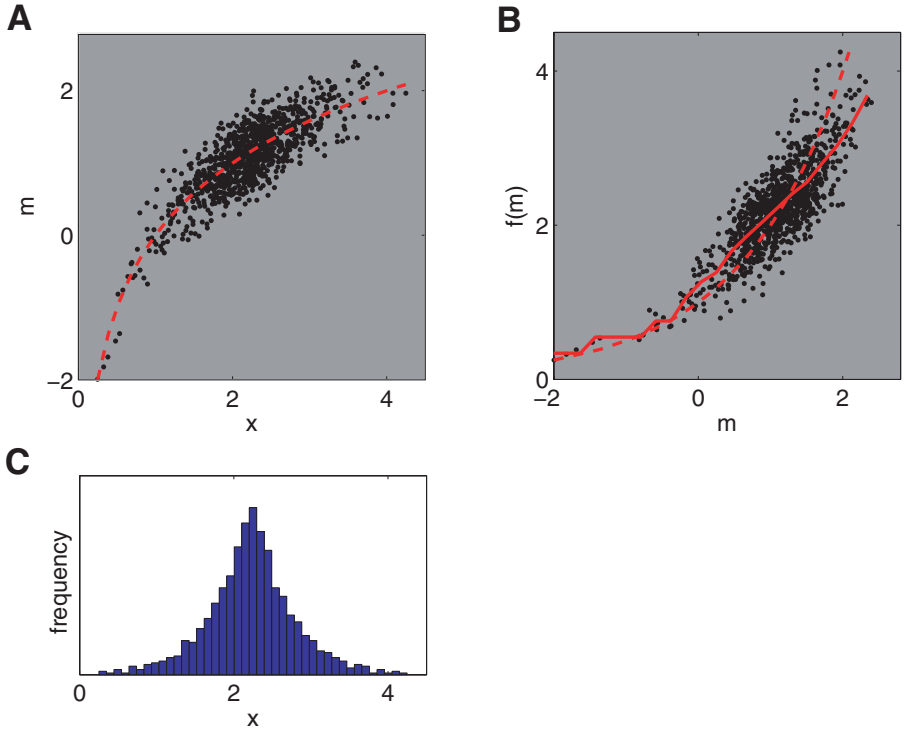


PLATE 49 Regression formulation of the optimal estimation problem, illustrated for a one-dimensional signal and measurement. (*A*) The measurement process (also known as the *encoding* process). We assume a set of data pairs (plotted points), $\{x_n, m_n\}$, indexed by $n \in [1, 2, \ldots, N]$, representing true signal values and associated noisy measurements. The dashed line indicates the average measurement as a function of the true signal value. (*B*) The estimation (or *decoding*) process. The estimator $f(m)$ maps measurements back to estimated signal values. The optimal estimator (solid line) does this so as to minimize a specified loss function. Note that this need not be (and is generally not) the inverse of the average measurement function (dashed line). Note also that the optimal estimator will depend on the signal values that are included in the data set, which are summarized by the histogram shown in panel *C*. (See figure 36.1.)

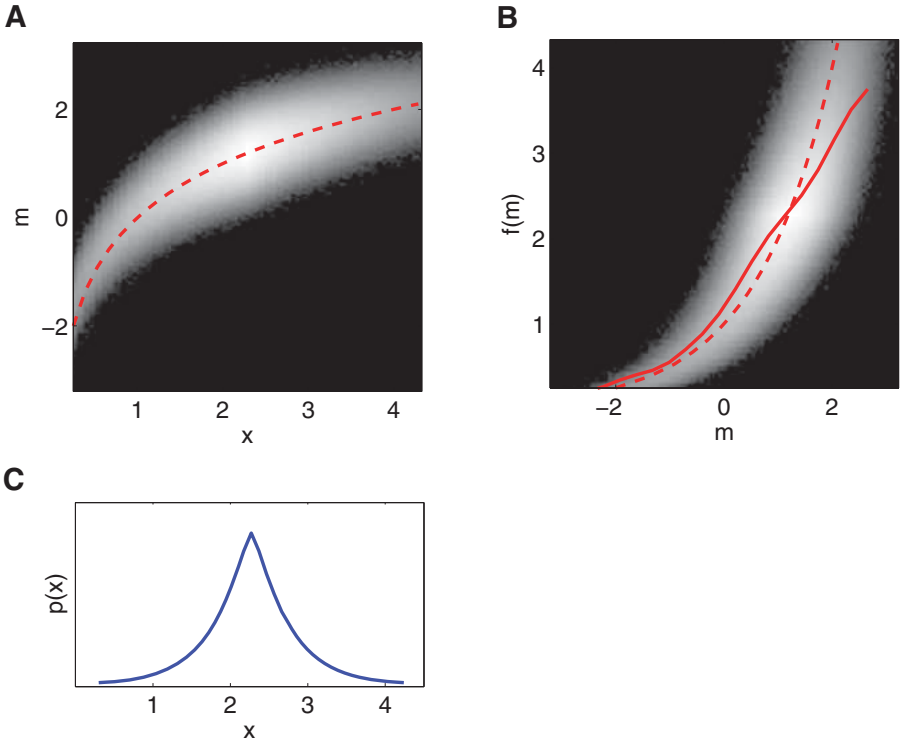


PLATE 50 Bayesian formulation of the optimal estimation problem. (*A*) The measurement density, $P(m \mid x)$, shown as a gray-scale image, where intensity indicates log probability. The dashed line indicates the mean of the density as a function of x. (*B*) The posterior density, $P(x \mid m)$. The solid line indicates the mean of the density, and the dashed line indicates the (inverted) mean of the measurement density in panel *A*. (*C*) the prior density, $P(x)$. (See figure 36.2.)

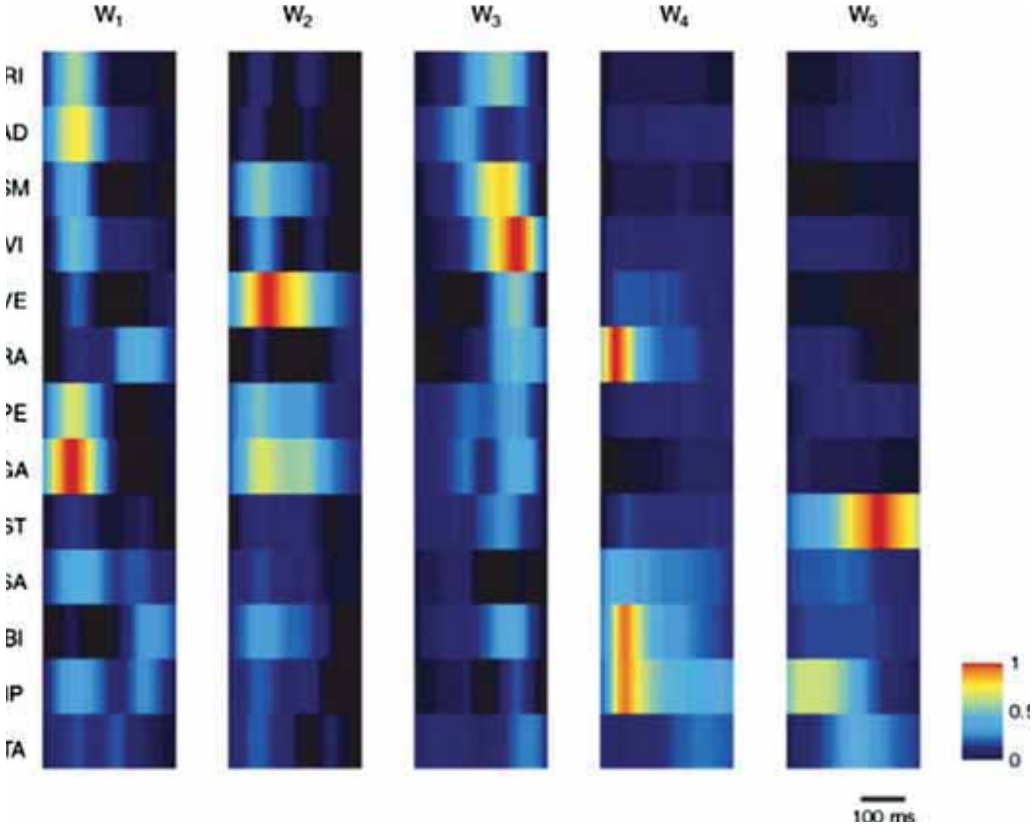


PLATE 51 Time-varying muscle synergies extracted from jumping, swimming, and walking muscle patterns in three frogs. Each synergy (columns W_1 to W_5) represents the activation time course (in color code) of 13 muscles over 30 samples (total duration: 300 ms) normalized to the maximum sample of each muscle. Abbreviations: RI, rectus internus; AD, adductor magnus; SM, semimembranosus; VI, the knee extensor vastus internus; VE, vastus externus; RA, rectus anterior; PE, the ankle extensors peroneus; GA, gastrocnemius; ST, mainly semitendinosus; SA, semitendinosus; BI, biceps; IP, iliopsoas; TA, tibialis anterior. (From Bizzi, E., Cheung, V. C., d'Avella, A., Saltiel, P., Tresch, M. C., 2008, Combining modules for movement, *Brain Res. Rev.*, *57*, 125–133.) (See figure 37.2.)

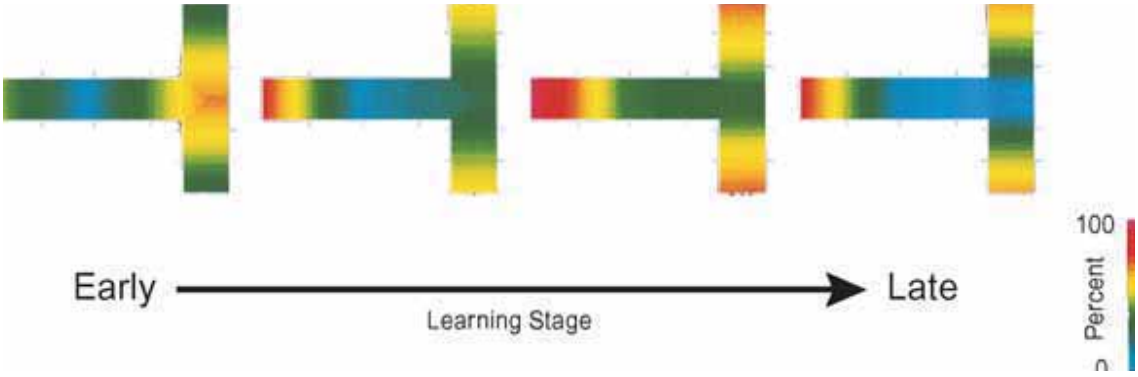


PLATE 52 Event-related ensemble activity of neurons in the dorsolateral striatum of rats recorded during the acquisition and performance of an auditory conditional turning task in a T-maze. Color plots at bottom illustrate schematically the gradual changes in the response profiles of the striatal neurons during the course of behavioral learning. (Modified from Jog et al., 1999; Graybiel and Kubota, 2003.) (See figure 39.11.)

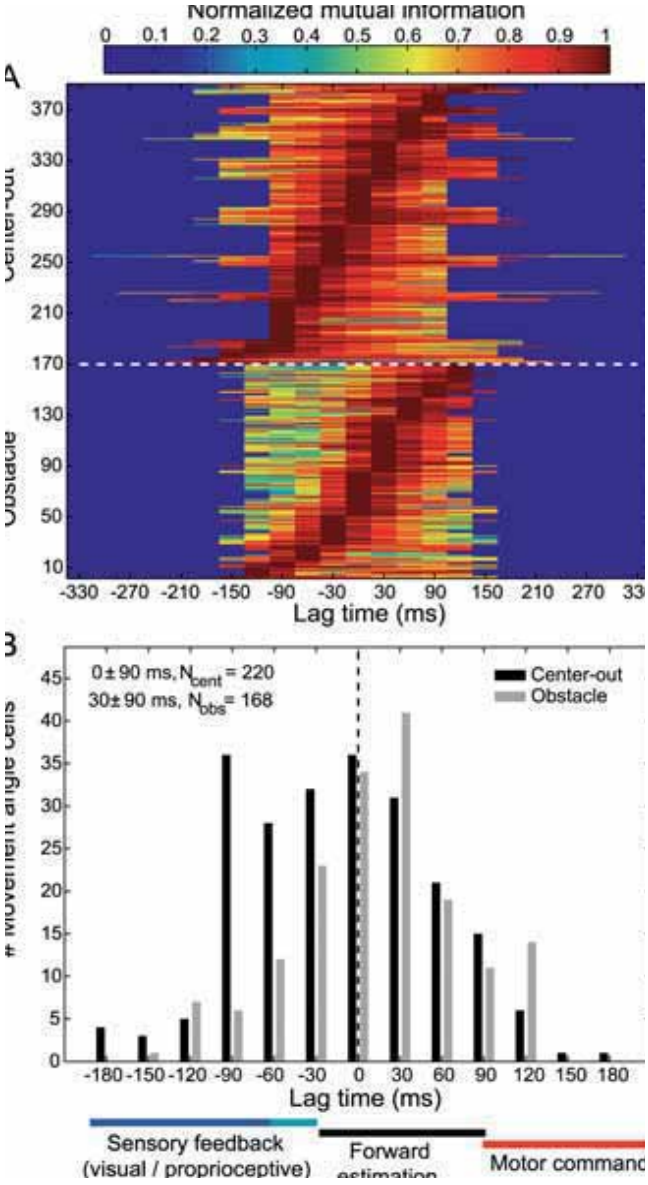


PLATE 53 Population temporal encoding results. (*A*) Population TEFs plotted for all movement angle neurons showing cell-normalized mutual information as a function of lag time. (*B*) Histogram summarizing the OLTs for movement angle neurons for both center-out and obstacle tasks (summary statistic in upper-left corner: median ± interquartile range). Many of these neuron's OLTs were consistent with a forward estimate of the state of the movement angle, which did not directly reflect delayed sensory feedback to PPC, nor were they compatible with outgoing motor commands from PPC. (Reprinted with permission from Mulliken, Musallam, & Andersen, 2008.) (See figure 41.3.)

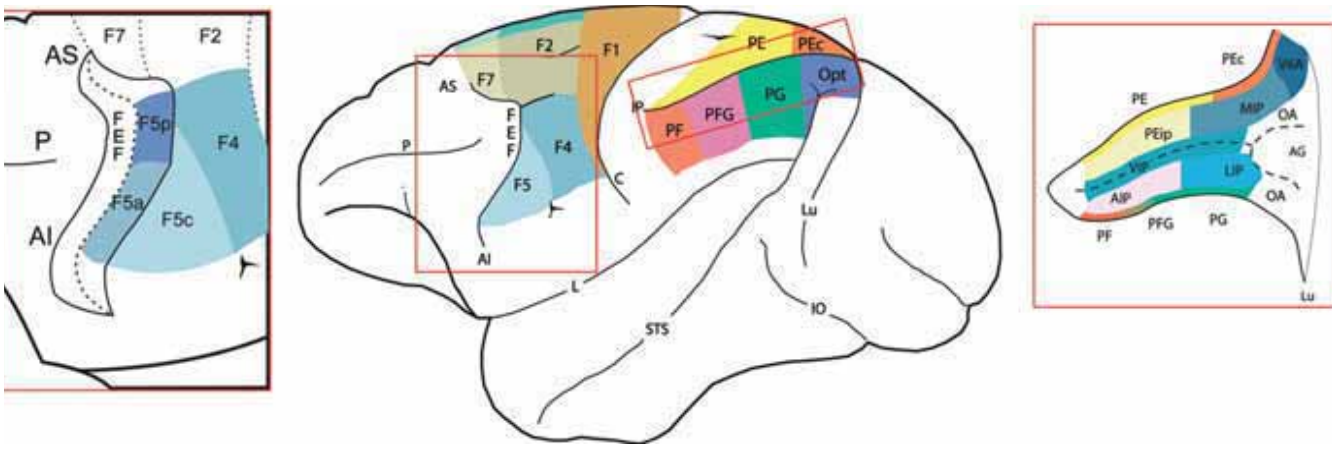


Plate 54 Lateral view of the monkey brain showing the parcellation of the motor and the posterior parietal cortex. The areas located within the arcuate and the intraparietal sulcus are shown in an unfolded view of these sulci in the left and right parts of the figure, respectively. For the nomenclature and definition, see Rizzolatti, Luppinon, and Matelli (1998), Nelissen et al. (2005), and Gregoriou et al. (2006). Abbreviations: AI, inferior arcuate sulcus; AS, superior arcuate sulcus; C, central sulcus; FEF, frontal eye-fields; IP, intraparietal sulcus; IO, inferior occipital sulcus; L, lateral fissure; Lu, lunate sulcus; P, principal sulcus; STS, superior temporal sulcus. (See figure 43.1.)

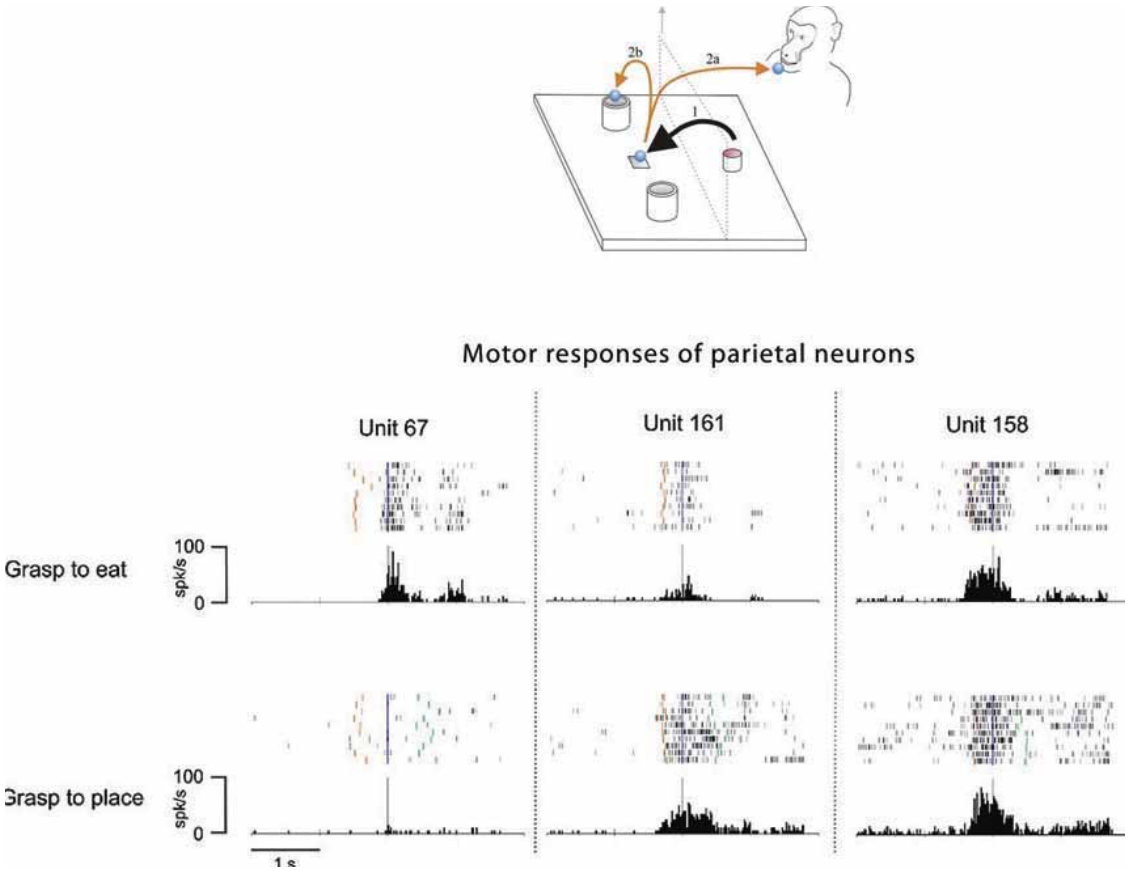


Plate 55 Examples of the activity of parietal motor neurons during execution of two different actions. (*A*) Apparatus and paradigm used for the motor task. In one condition (grasping for eating), the monkey reached for and grasped a piece of food located on a plane in front of it (1) and brought the food to its mouth (2a). In another condition (grasping for placing), the monkey reached for and grasped an object located in front of it (1) and placed the object into a container (2b). In the first condition, the monkey ate the food that it had brought to the mouth; in the second condition, the monkey was rewarded after correct accomplishment of the task. (*B*) Activity of three IPL neurons during grasping in the two experimental conditions. Unit 67 discharges were stronger during grasping to eat than during grasping to place, Unit 161 discharges were stronger during grasping to place. Unit 158 did not show any difference in discharge between the two conditions. Rasters and histograms are aligned with the moment when the monkey touched the object or food to be grasped. Red bars: Monkey releases the hand from the starting position. Green bars: Monkey touches the container. Abscissa: Time, bin = 20 ms; Ordinate: Discharge frequency in spikes per second. (Modified from Fogassi et al., 2005.) (See figure 43.3.)

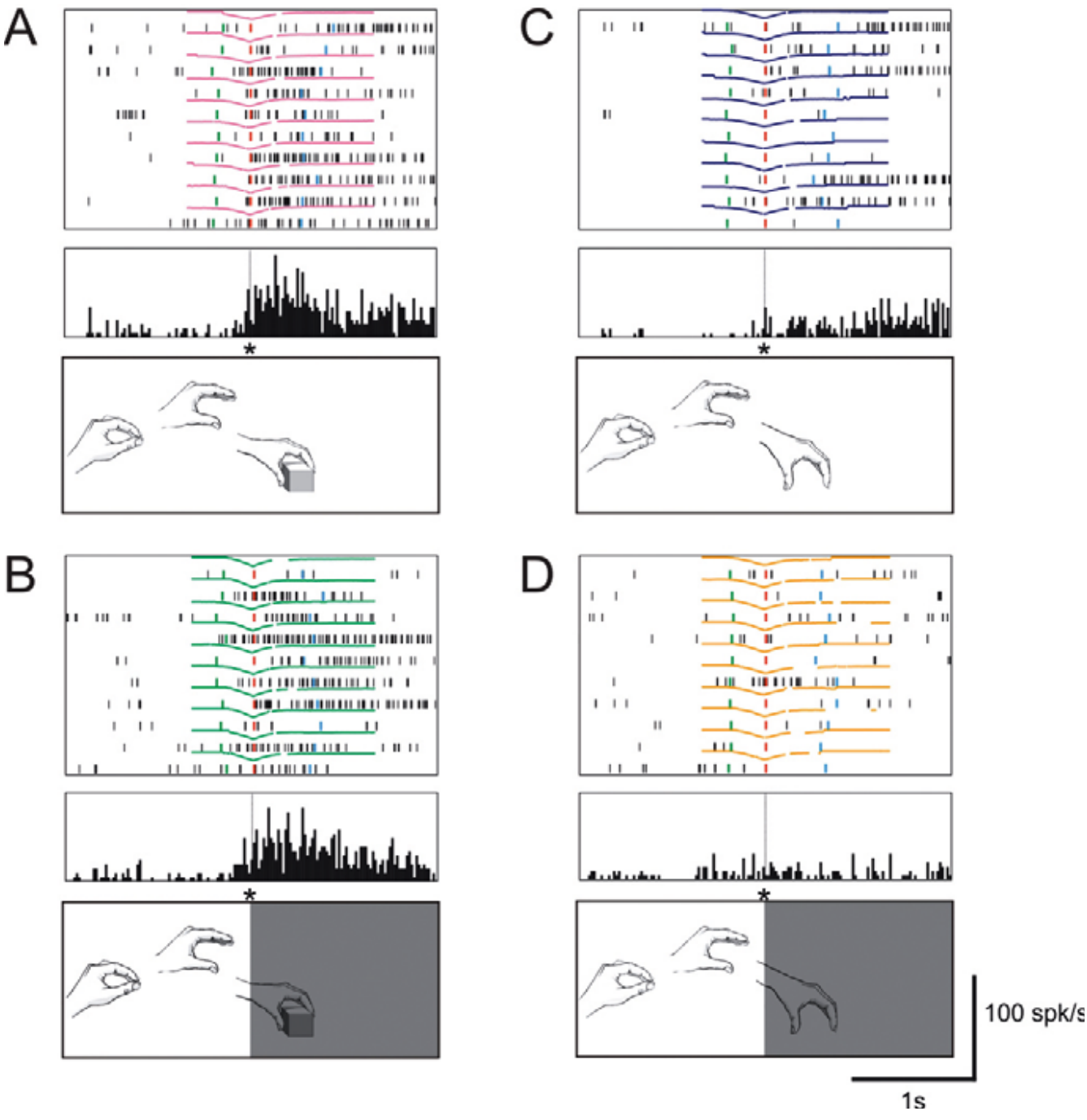


PLATE 56 Example of a mirror neuron responding during observation of grasping in both full vision and "hidden" condition. (*A* and *C*): Observation of goal-directed or mimed grasping, respectively, in full vision. (*B* and *D*): Observation of goal-directed or mimed grasping, respectively, in the hidden condition. In every panel, from top to bottom, rasters and histogram and the schematic drawing of the experimenter motor act are shown. The gray frame in conditions *B* and *D* represents a screen interleaved between the monkey and the experimenter hand in the two hidden conditions. The asterisk indicates the location of a stationary marker that was attached at the level of the crossing point where the experimenter's hand disappeared behind the screen in the hidden conditions. The colored line above each raster represents the kinematics of the experimenter's hand movement; the downward deflection of the line means that the hand is approaching the stationary marker (the minimum corresponding to the moment in which the hand is closest to the marker). Histograms bin width = 20 ms. (Modified from Umiltà et al., 2001.) (See figure 43.5.)

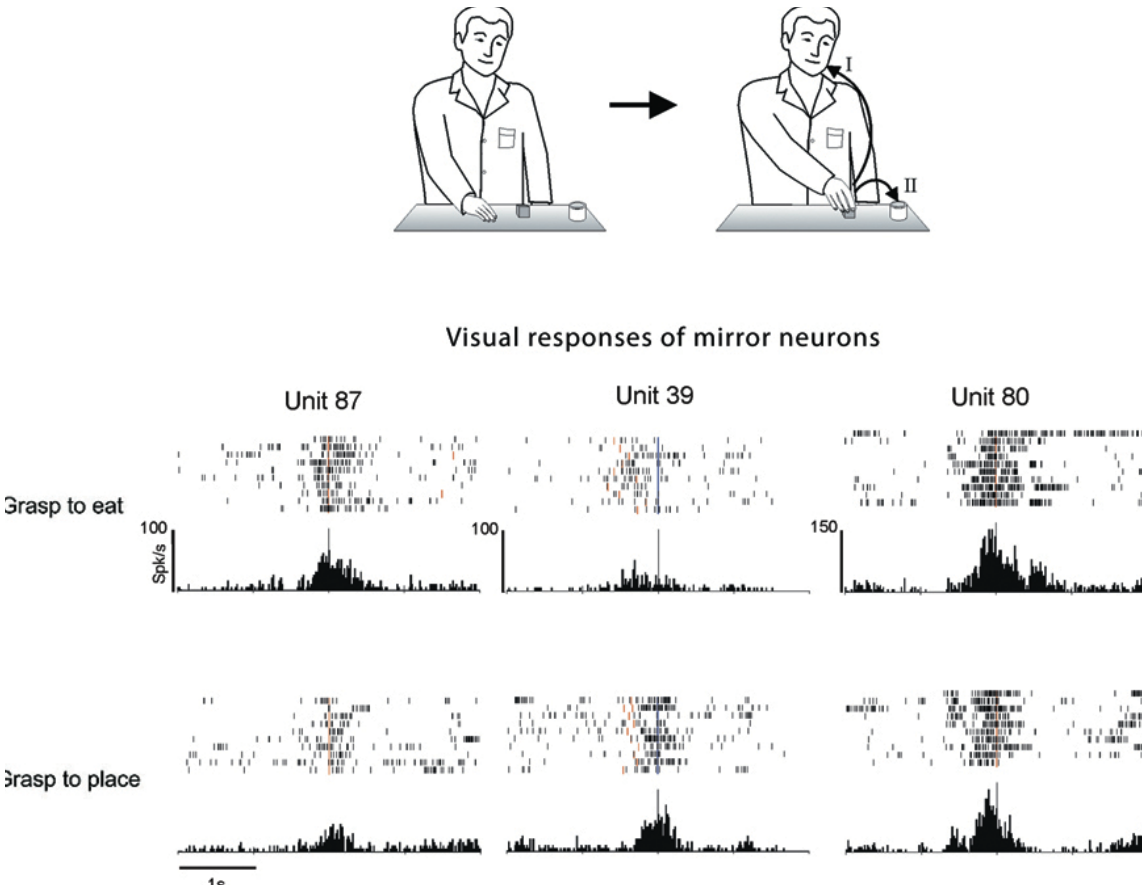


PLATE 57 Examples of visual responses of IPL mirror neurons during the observation of grasping-to-eat and grasping-to-place conditions performed by an experimenter. (*A*) The paradigm is similar to that used for the motor task shown in figure 43.3, but in this case, the two conditions are performed by the experimenter in front of the monkey, which is simply observing the scene. (*B*) Activity of three mirror neurons during observation of grasping in the two experimental conditions. Unit 87 discharges are stronger during observation of grasping to eat than during observation of grasping to place, Unit 39 discharges are stronger during observation of grasping to place. Unit 80 did not show any difference in discharge between the two conditions. Rasters and histograms are aligned with the moment when the experimenter touched the object or food to be grasped. (Modified from Fogassi et al., 2005.) (See figure 43.6.)

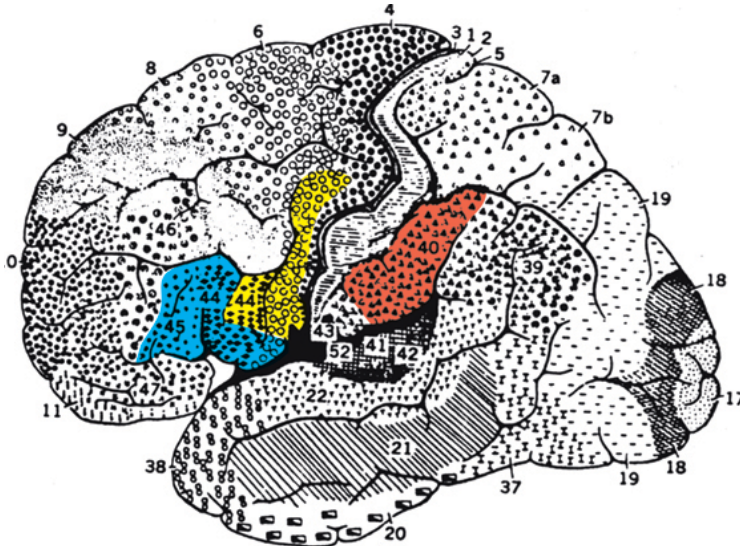


PLATE 58 Lateral view of the human cortex showing the frontal (yellow and blue) and parietal (red) regions constituting the core of the grasping mirror neuron system in humans. Numbers and symbols indicate the different cytoarchitectonic areas according to the parcellation of Brodmann (1909). (See figure 43.7.)

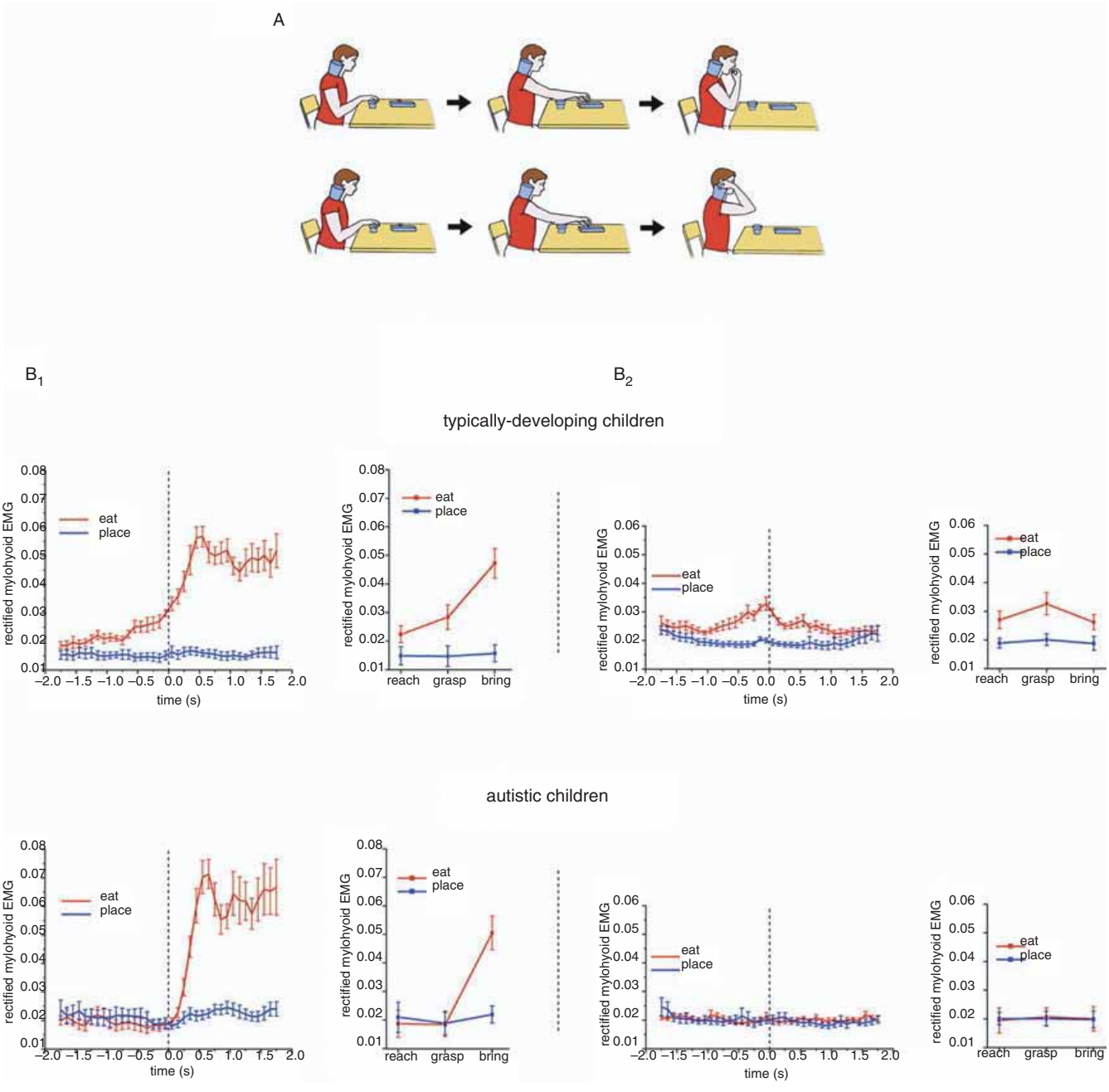


PLATE 59 Differential activation of a mouth-opening muscle during execution and observation of two actions in typically developing and autistic children. (*A*) Schematic representation of the two actions executed and observed by the two groups of subjects. *Upper part*: The individual reaches for and grasps a piece of food located on a touch-sensitive plate, brings it to the mouth, and eats it. *Lower part*: The individual reaches for and grasps a piece of paper located on the same plate and puts it into a container placed on the shoulder. (B_1) *Left*: Time course of the EMG activity of the mylohyoid muscle during execution of grasping for eating (red) and grasping for placing (blue). Vertical bars indicate the standard error. The curves are aligned (dashed vertical line) with the moment in which the object is lifted from the touch-sensitive plate. *Right*: Mean EMG activity of the same muscle in three epochs of the two actions. Vertical bars indicate 95% confidence intervals. (B_2) *Left*: Time course of the EMG activity of the mylohyoid muscle during observation of grasping for eating (red) and grasping for placing (blue). Other conventions as in B_1. *Right*: Mean EMG activity of the same muscle in three epochs of the two observed actions. Other conventions as in B_1. (Modified from Cattaneo et al., 2008.) (See figure 43.8.)

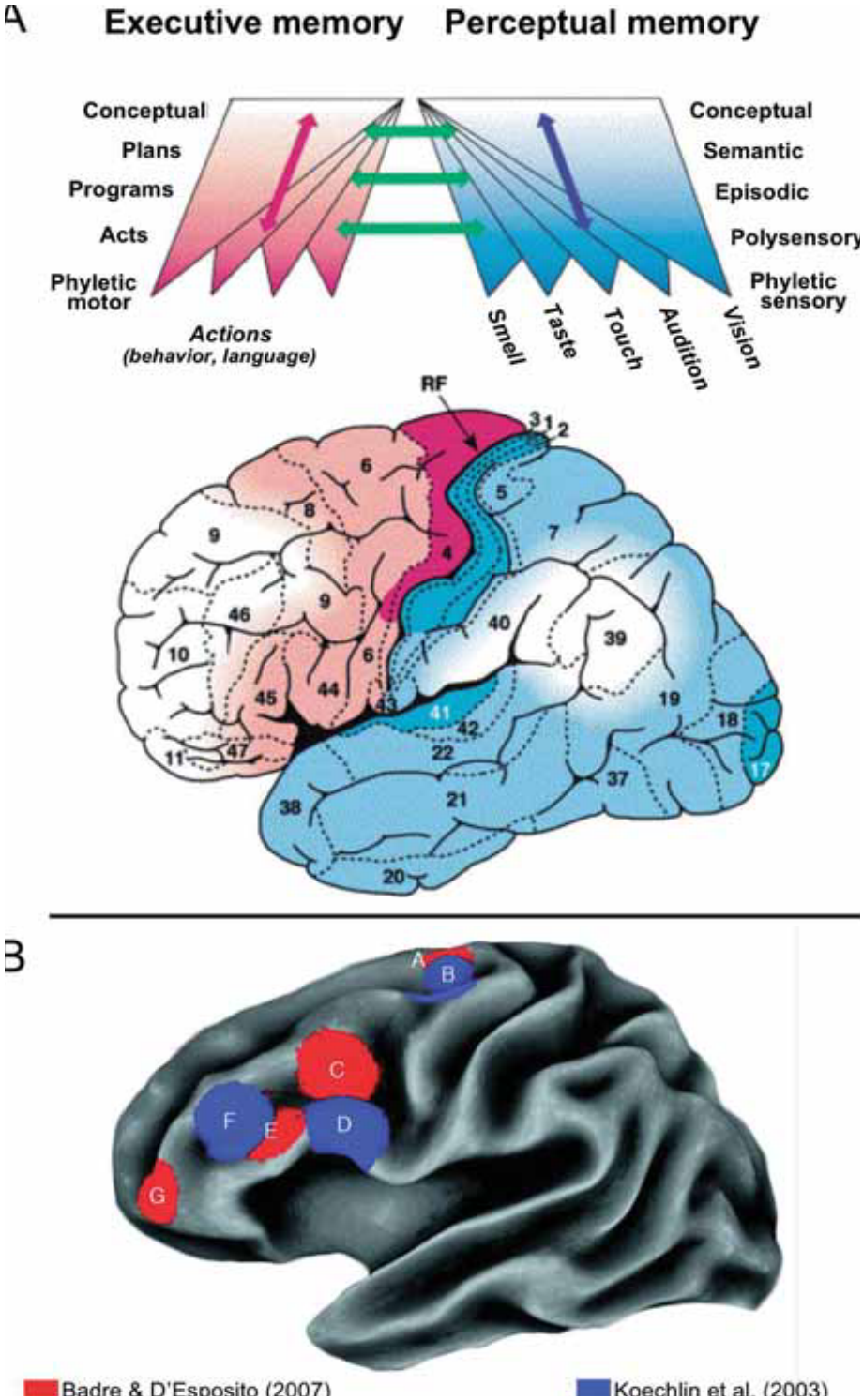


PLATE 60 Hierarchical organization of cognitive representations in lateral cortex. (*A*) Schema of two hierarchies of cortical memory, executive memory, and perceptual memory, and the distribution of these hierarchies in frontal and posterior cortical regions, respectively. In frontal cortex, representations that are "higher" in the processing hierarchy are mapped to more rostral regions, and "lower"-level representations are mapped to more caudal regions. (Reprinted from J. M. Fuster, 2001, The prefrontal cortex—An update: Time is of the essence, *Neuron*, *30*, 319–333. Copyright 2001, with permission from Elsevier.) (*B*) Neuroimaging data providing evidence for representational hierarchies in frontal cortex. Spheres from Badre and D'Esposito (2007) (red) reflect foci of activation with experimental manipulations at different levels of representation: A, the response level; C, the feature level; E, the dimension level; G, the context level. Spheres from Koechlin, Ody, and Kouneiher (2003) (blue) reflect foci of activation with manipulations of different levels of control: B, sensory control; D, contextual control; F, episodic control. (Adapted with permission from D. Badre & M. D'Esposito, 2007, Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex, *J. Cogn. Neurosci.*, *19*, 2082–2099. Copyright 2007, with permission from the MIT Press.) (See figure 48.3.)

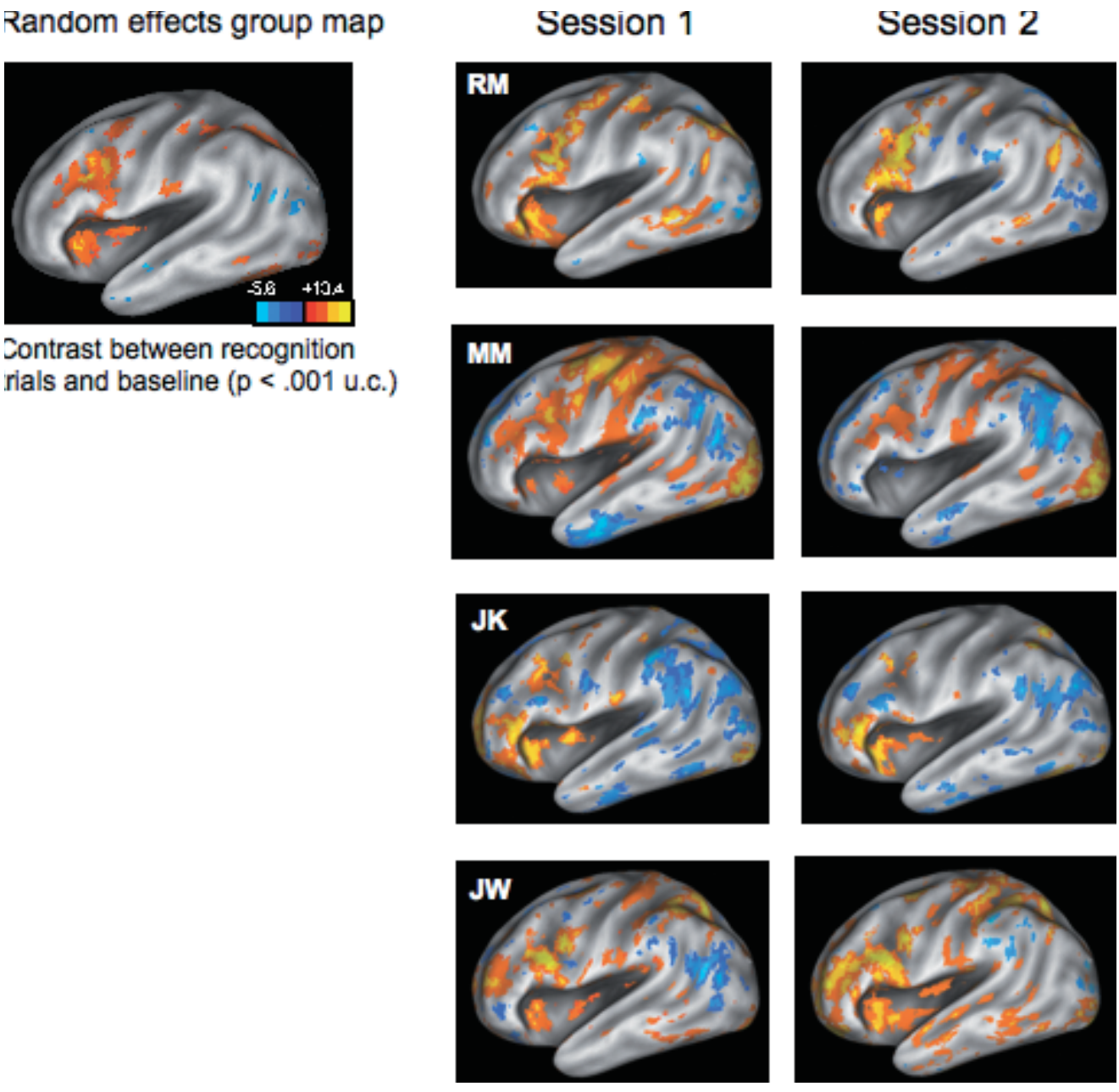


PLATE 61 The left hemisphere patterns of activations and deactivations during an episodic retrieval task for the group and for 4 of the 14 individuals that made up the group. The random-effects group map and the individual *t*-maps are a contrast between the retrieval condition and baseline, statistically thresholded at $p < .001$ uncorrected for multiple comparisons. The thresholding was done for visualization purposes only. Displayed for each individual is the *t*-map for the first and second sessions. The images are representative of the high degree of variability between individuals, yet the relative stability in the individual patterns of activity over time. (From Miller et al., in press.) (See figure 50.1.)

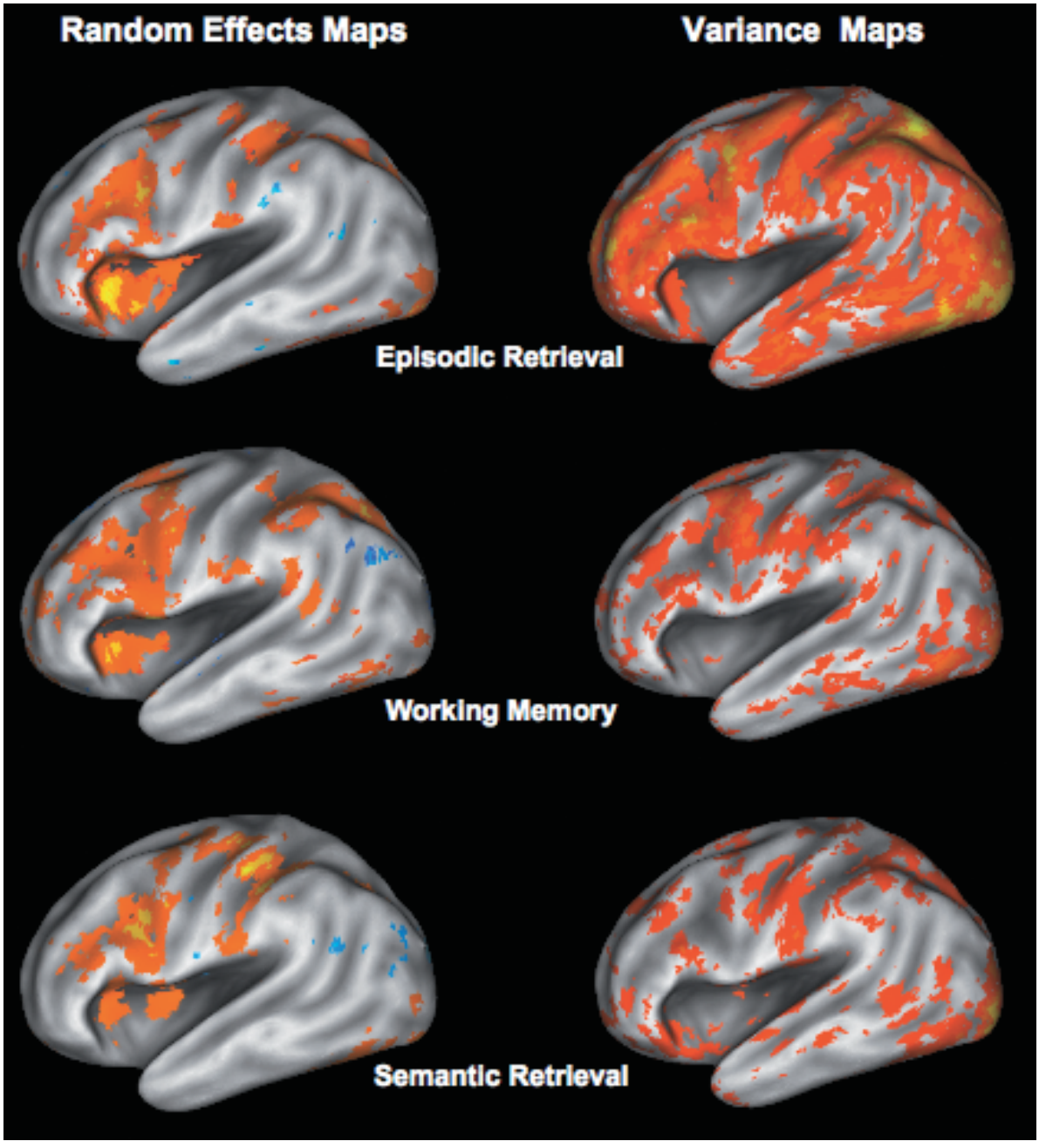


Plate 62 A comparison of random-effects group maps and variance maps across three memory tasks. The random-effects maps are a statistically thresholded ($p < .001$ uncorrected for multiple comparisons) representation of the common areas of brain activity across 14 individuals. The variance maps to the right display the standard deviations across the 14 individuals at each voxel above a threshold of 2 standard deviations. As the variance maps indicate, individuals variably engaged much wider regions of the cortex during episodic retrieval than during semantic retrieval or working memory. (See figure 50.2.)

Plate 64 The dual-stream model of speech processing. (*A*) Schematic diagram of the dual-stream model. The earliest stage of cortical speech processing involves some form of spectrotemporal analysis, which is carried out in auditory cortices bilaterally in the supratemporal plane. These spectrotemporal computations appear to differ between the two hemispheres. Phonological-level processing and representation involves the middle to posterior portions of superior temporal sulcus (STS) bilaterally, although there may be a weak left-hemisphere bias at this level of processing. Subsequently, the system diverges into two broad streams, a dorsal pathway (blue) that maps sensory or phonological representations onto articulatory motor representations, and a ventral pathway (red) that maps sensory or phonological representations onto lexical-conceptual representations. (*B*) Approximate anatomical locations of the dual-stream model components, specified as precisely as available evidence allows. Regions shaded green depict areas on the dorsal surface of the STG that are hypothesized to be involved in spectrotemporal analysis. Regions shaded yellow in the posterior half of the STS are implicated in phonological-level processes. Regions shaded red

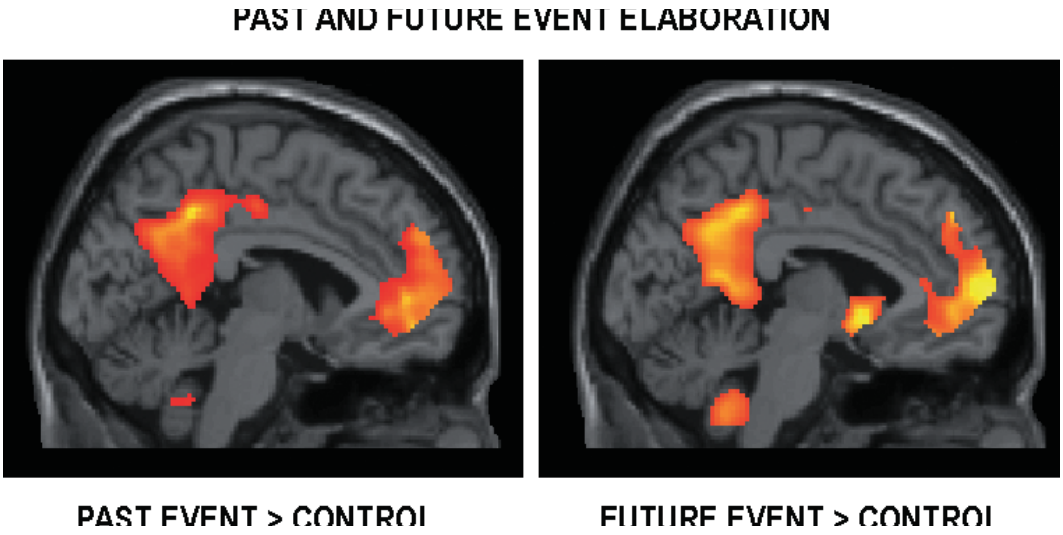
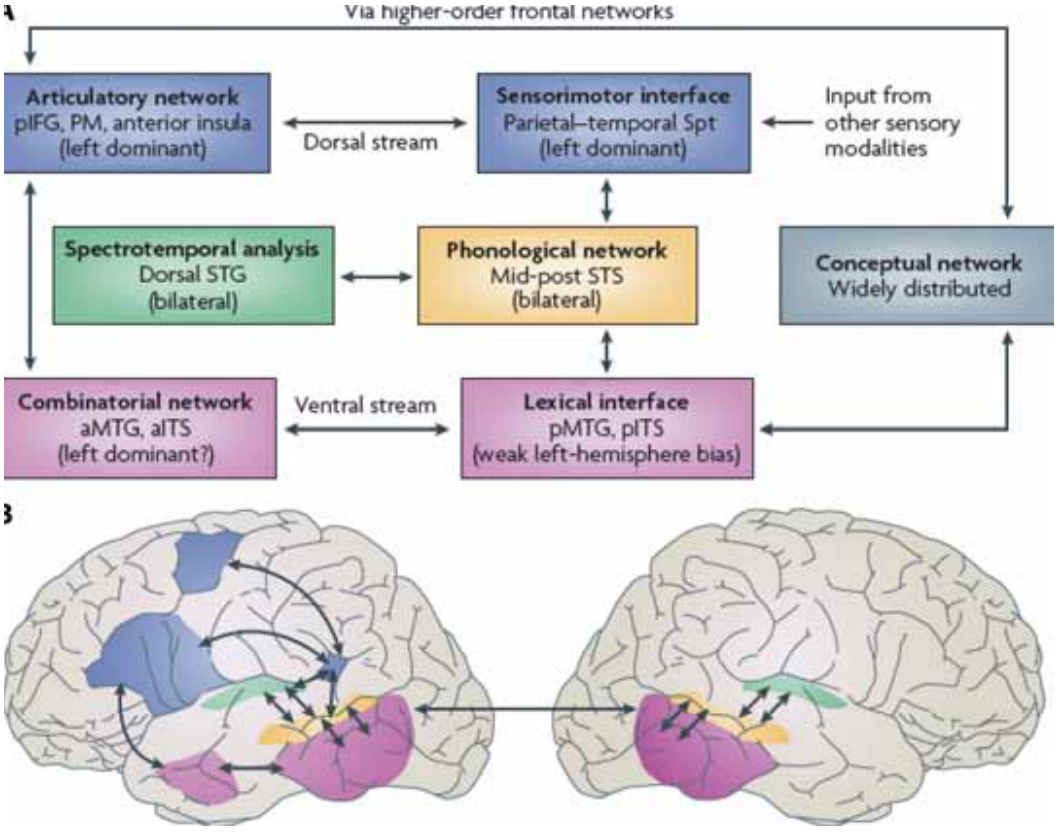


PLATE 63 Sagittal slice ($x = -4$) illustrating the striking commonalities in the medial left prefrontal and parietal regions engaged when remembering the past (*left panel*) and imagining the future (*right panel*). These marked similarities of activation were also evident in areas of the medial temporal lobe (left hippocampus, bilateral parahippocampal gyrus) and lateral cortex (left temporal pole and left bilateral inferior parietal cortex). This extensive pattern of common activity was not present during the construction of past and future events; it only emerged during the elaboration of these events (shown here, relative to the elaboration phase of a semantic and an imagery control task; significant at $p < .001$, uncorrected; shown at $p < .005$, uncorrected.) (Originally published in Addis, Wong, & Schacter, 2007.) (See figure 51.1.)



represent the ventral stream, which is bilaterally organized with a weak left-hemisphere bias. The more posterior regions of the ventral stream, the posterior middle and inferior portions of the temporal lobes, correspond to the *lexical interface*, which links phonological and semantic information, whereas the more anterior locations correspond to the hypothesized *combinatorial network*. Regions shaded blue represent the dorsal stream, which is strongly left-dominant. The posterior region of the dorsal stream corresponds to an area in the Sylvian fissure at the parietal-temporal boundary (area Spt), which is hypothesized to be a *sensorimotor interface*, whereas the more anterior locations in the frontal lobe, likely involving Broca's region and a more dorsal premotor site, correspond to portions of the articulatory network. (Figure reproduced from Hickok & Poeppel, 2007.) (See figure 52.2.)

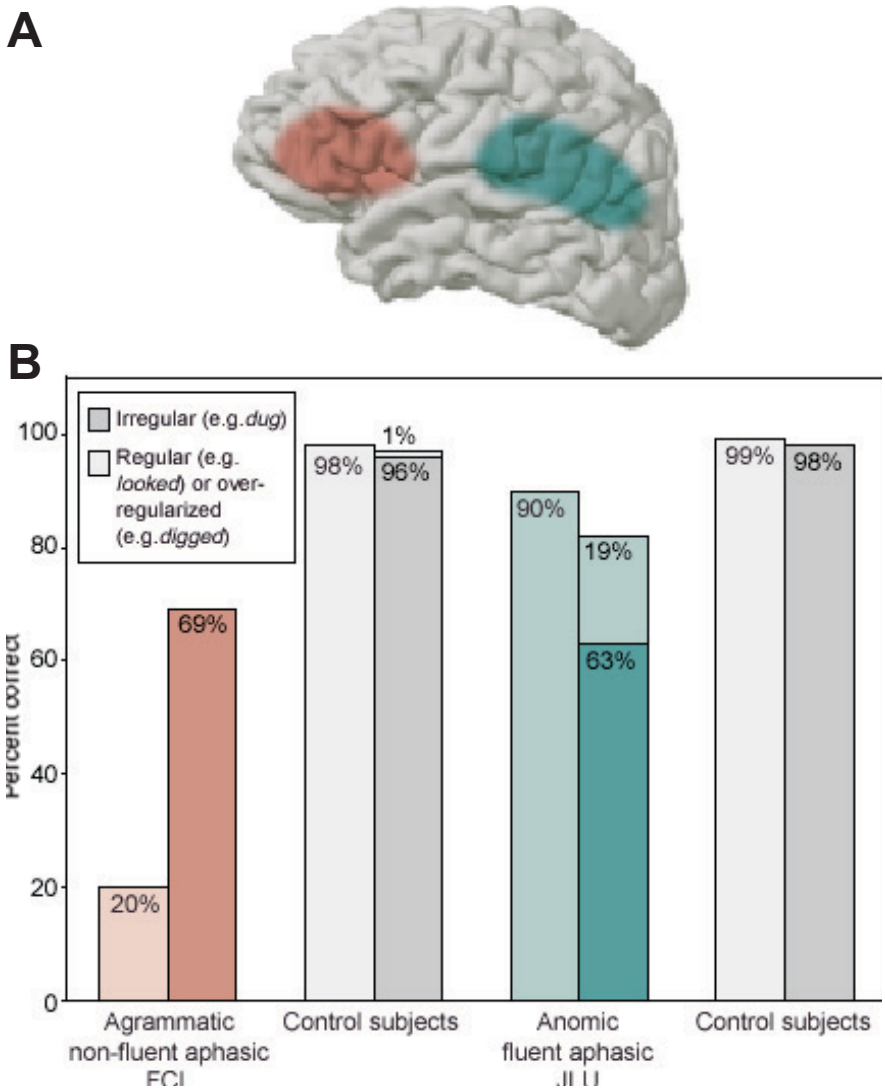


PLATE 65 A neuropsychological dissociation in processing regular and irregular verb forms. (*A*) The approximate lesion sites of patient FCL (red area, left anterior perisylvian regions), who had symptoms of agrammatism, and patient JLU (green area, left temporoparietal region), who had symptoms of anomia. (*B*) Results of verb inflection tests showed that the agrammatic patient had more trouble inflecting regular verbs (lighter bars) than irregular verbs (darker bars), whereas the anomic patient had more trouble inflecting irregular verbs—and overapplied the regular suffix to many of the irregulars (light green bar on top of dark green bar). The performance of age- and education-matched control subjects is shown in the gray bars. (Reprinted from Pinker & Ullman, 2002.) (See figure 53.1.)

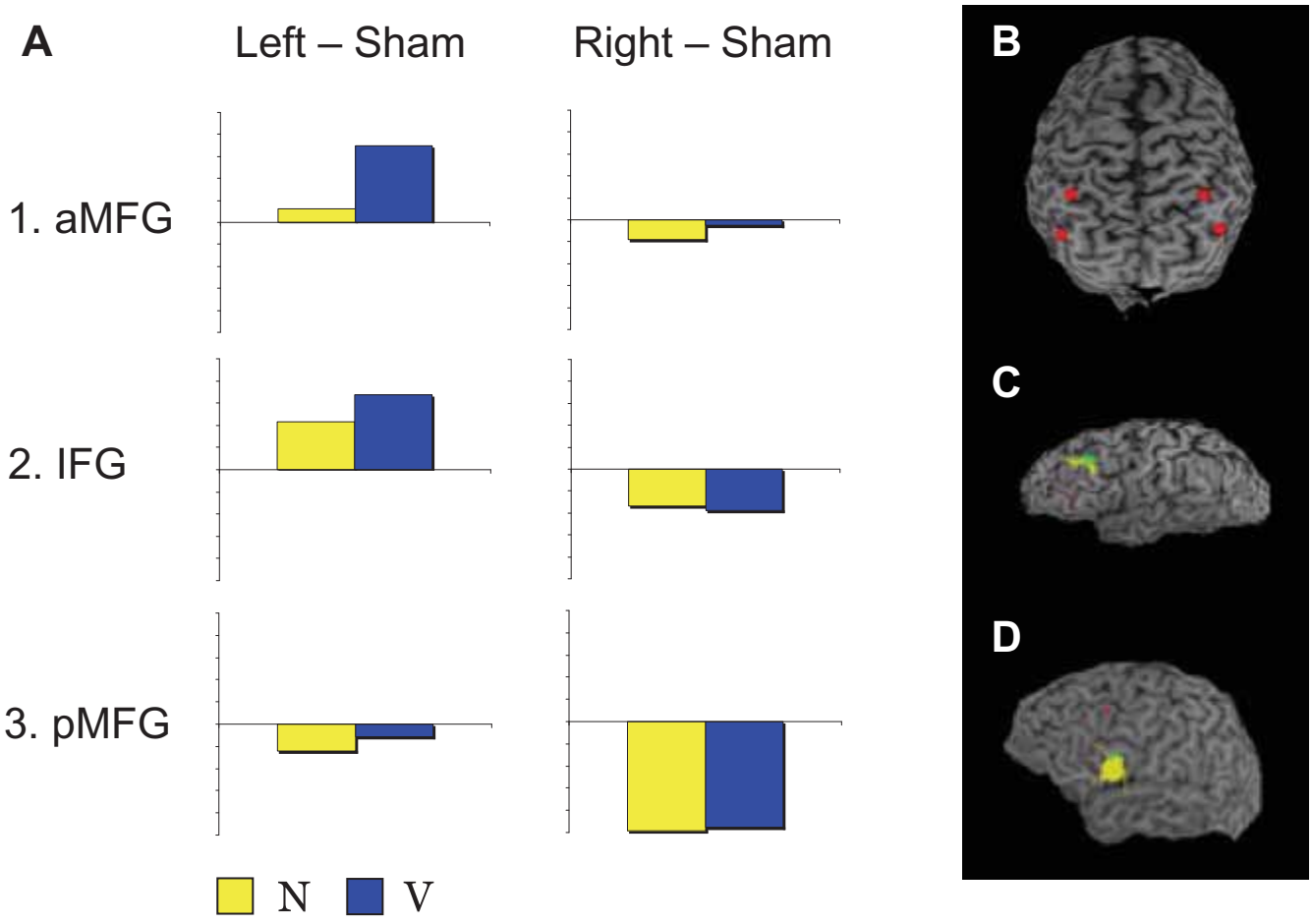


Plate 66 Results of the rTMS experiment reported by Cappelletti et al. (2008), showing a selective disruption in verb processing following stimulation to the left anterior frontal gyrus. (*A*) The mean difference in reaction times to nouns and verbs with repetitive TMS compared to sham stimulation in three areas: the anterior middle frontal gyrus (aMFG), inferior frontal gyrus (IFG), and posterior middle frontal gyrus (pMFG). (*B*) The sites of stimulation to the IFG and pMFG. The remaining panels demonstrate the stereotactic application of TMS to the left pMFG (*C*) and left IFG (*D*). (Modified from Cappelletti, Fregni, Shapiro, Pascual-Leone, & Caramazza, 2008.) (See figure 53.2.)

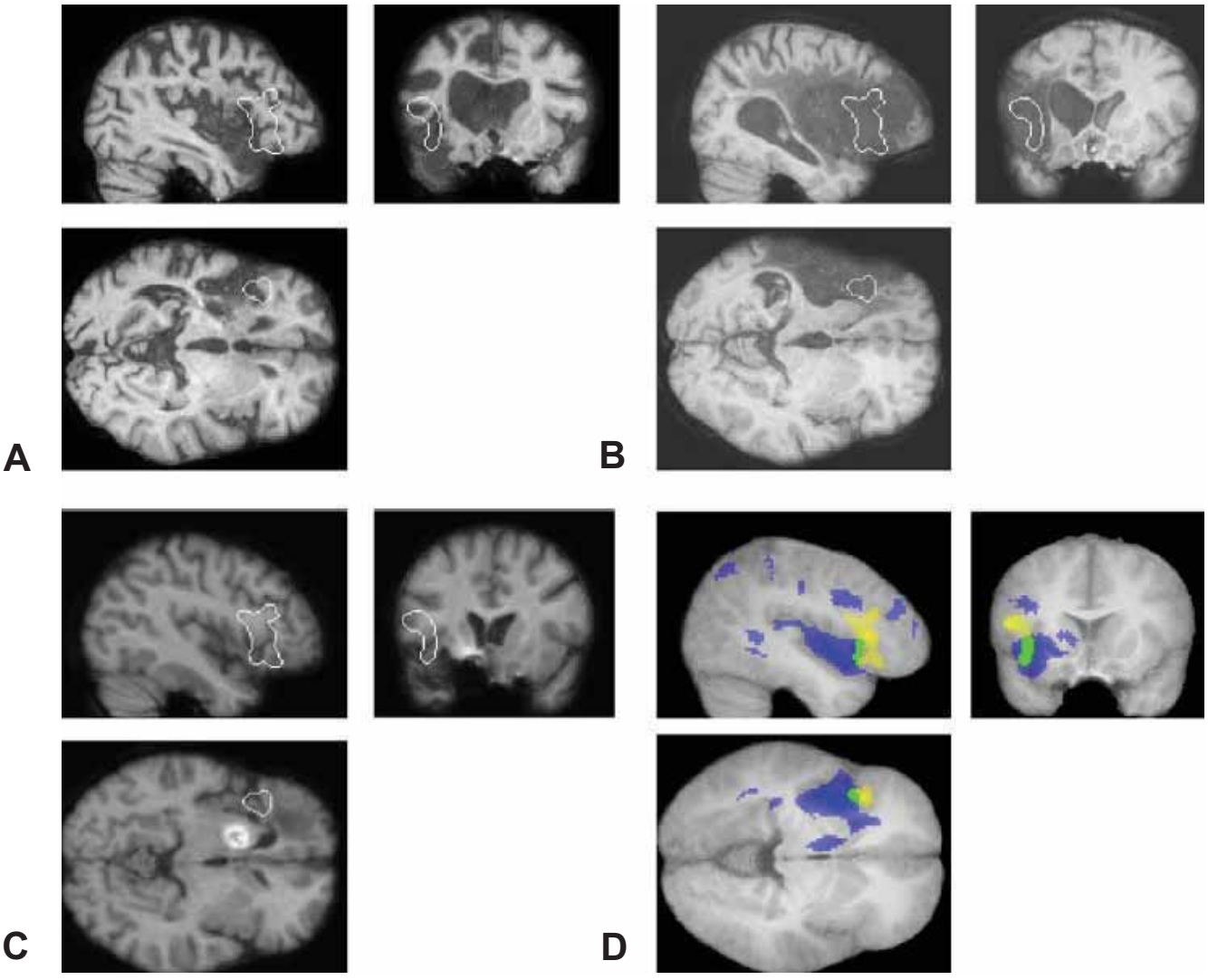


Plate 67 The area found by Tyler and colleagues (2004) to be more active for inflected verbs than inflected nouns in an fMRI semantic judgment paradigm, compared to the lesion sites of three aphasic patients with deficits in processing regularly inflected verb forms in a priming task. (*A–C*) T1-weighted MR images of three patients with an outline of the activation found in the verbs-nouns contrast superimposed on them. (*D*) A mean of the spatially normalized T1 images of the 12 subjects in the fMRI experiment overlaid with the lesion overlap of the three patients in *A–C*. Lesion overlap is shown in blue, the significant activation found in the verbs-nouns contrast is in yellow, and the overlap between common lesion volume of the three patients and the activation is in green. (Reprinted from Tyler, Bright, Fletcher, & Stamatakis, 2004.) (See figure 53.3.)

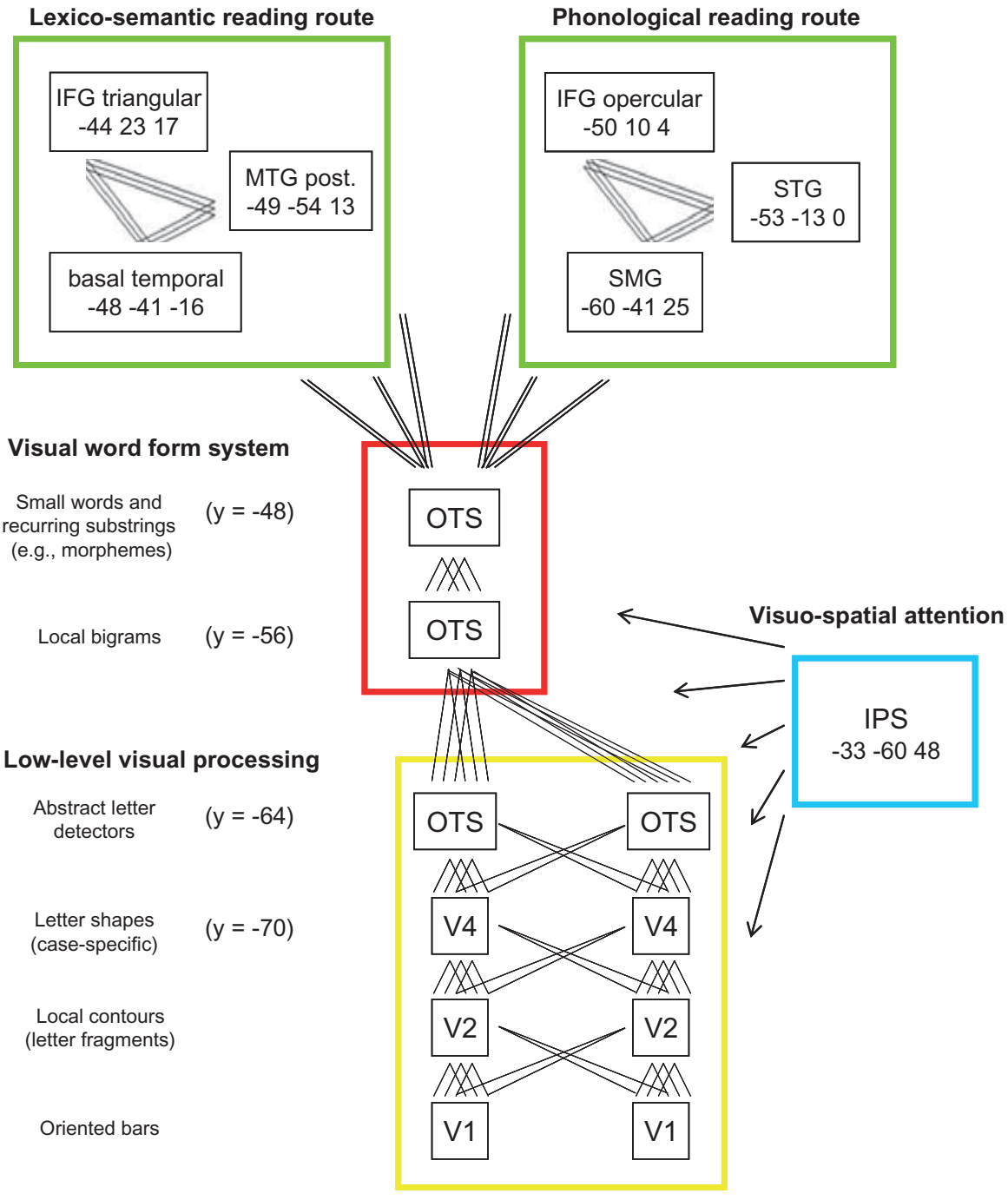


PLATE 68 Synthetic schema of the reading system, merging propositions from Dehaene, Cohen, Sigman, and Vinckier (2005) and Cohen and colleagues (2003). Low-level processing is achieved in each hemisphere for the contralateral half of the visual field (yellow). Information converges on the left-hemispheric Visual Word Form system, where an invariant representation of letter strings is computed (red). The dorsal visual stream exerts a top-down attentional control on the hierarchy of ventral areas (blue). The ventral visual system then feeds the lexicosemantic and phonological reading routes (green). The proposed normalized coordinates for the lexicosemantic and phonological reading routes are from a meta-analysis of 35 PET and fMRI studies (Jobard, Crivello, & Tzourio-Mazoyer, 2003), and the coordinates of the visuospatial attention system are from Gitelman et al. (1999). IFG: inferior frontal gyrus; MTG: middle temporal gyrus; SMG: supramarginal gyrus: OTS: occipitotemporal sulcus; IPS: intraparitetal sulcus. (See figure 54.1.)

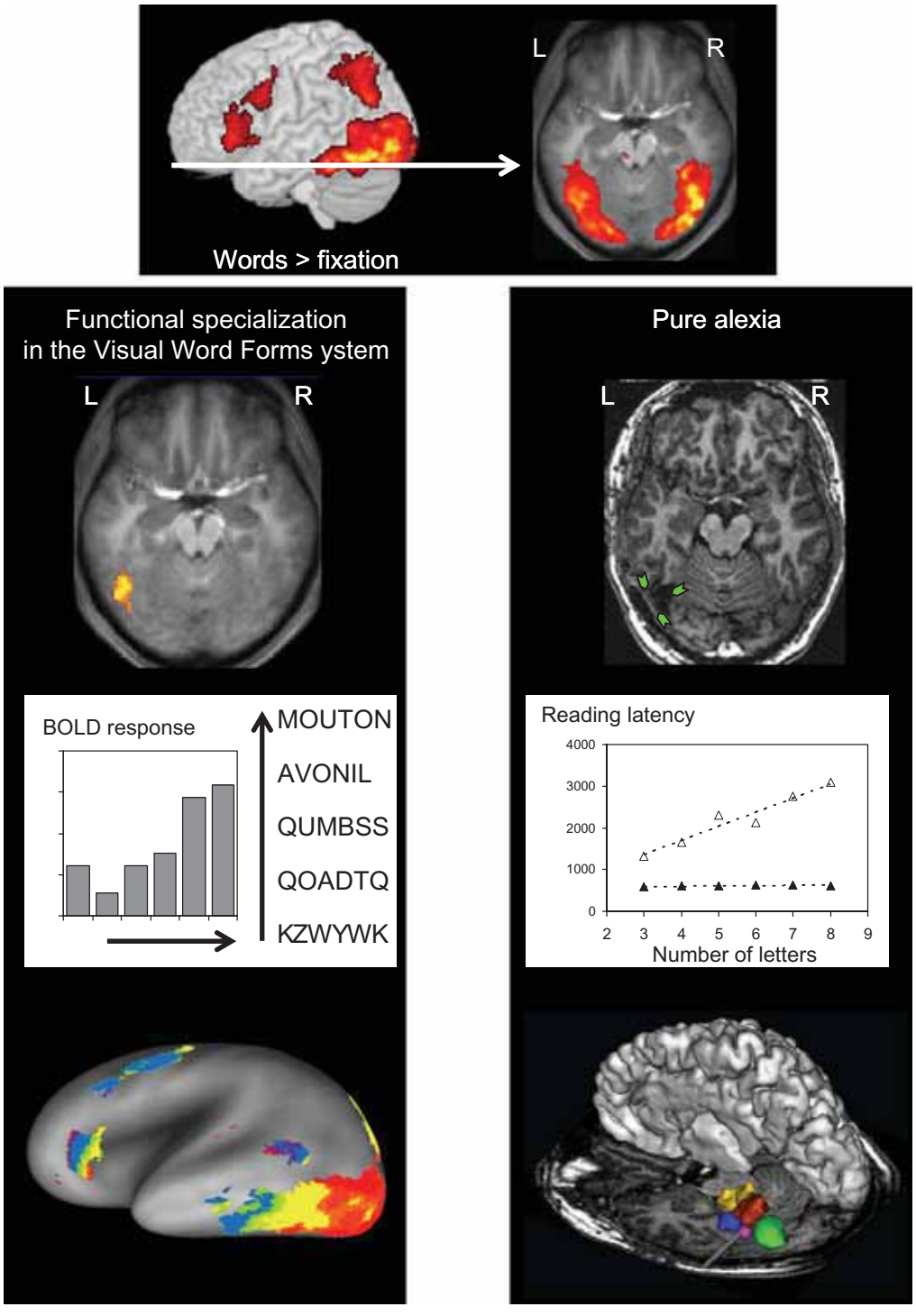


Plate 69 Word processing in the ventral pathway. *Top panel*: Activations induced by printed words relative to a fixation baseline in the left hemisphere (left) and in the bilateral ventral visual pathway (right). *Left panel*: The VWF system shows a linear increase of activation (top) by letter strings forming closer statistical approximations to orthographically legal strings (middle). This functional specialization increases progressively in more anterior regions within the VWF system (bottom). (Left panel adapted from Vinckier et al., 2007.) *Right panel*: Surgical lesion in the left ventral cortex responsible for pure alexia (top). Whereas before surgery word reading was fast and constant irrespective of word length, after surgery the patient showed slow letter-by-letter reading (middle). In the same patient, the 3D image shows the relative position of the VWFA (blue), of other category-dependent fMRI activation clusters before surgery, of the brain lesion (green), and of intracerebral electrodes (magenta). (Right panel adapted from Gaillard et al., 2006.) (See figure 54.2.)

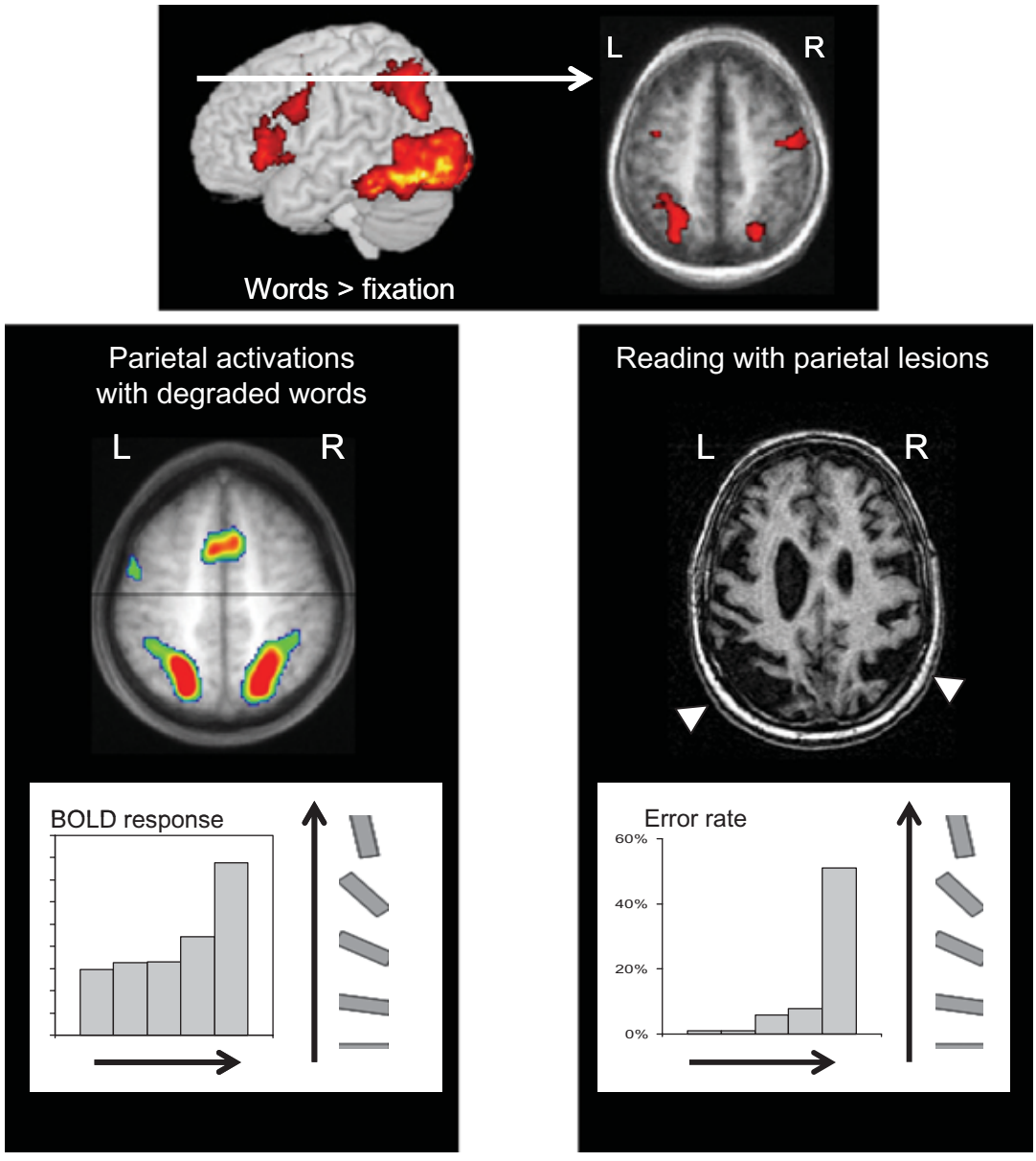


PLATE 70 Contribution of the dorsal pathway to word reading. *Top panel*: Activations induced by printed words relative to a fixation baseline in the left hemisphere (left) and in the bilateral dorsal visual pathway (right). *Left panel*: The bilateral intraparietal cortex shows a nonlinear increase of activation with word degradation, correlated with reaction times (top). For instance, activations increased steeply for words rotated by more than 45° (bottom). (Left panel adapted from Cohen, Dehaene, Vinckier, Jobert, & Montavont, 2008.) *Right panel*: In a patient with bilateral parietal atrophy and spared ventral cortex (top), there was a severe reading impairment above a similar threshold of rotation angle, demonstrating the role of parietal cortex whenever display degradation exceeds the range of invariance in the ventral cortex. (Right panel adapted from Vinckier et al., 2006.) (See figure 54.3.)

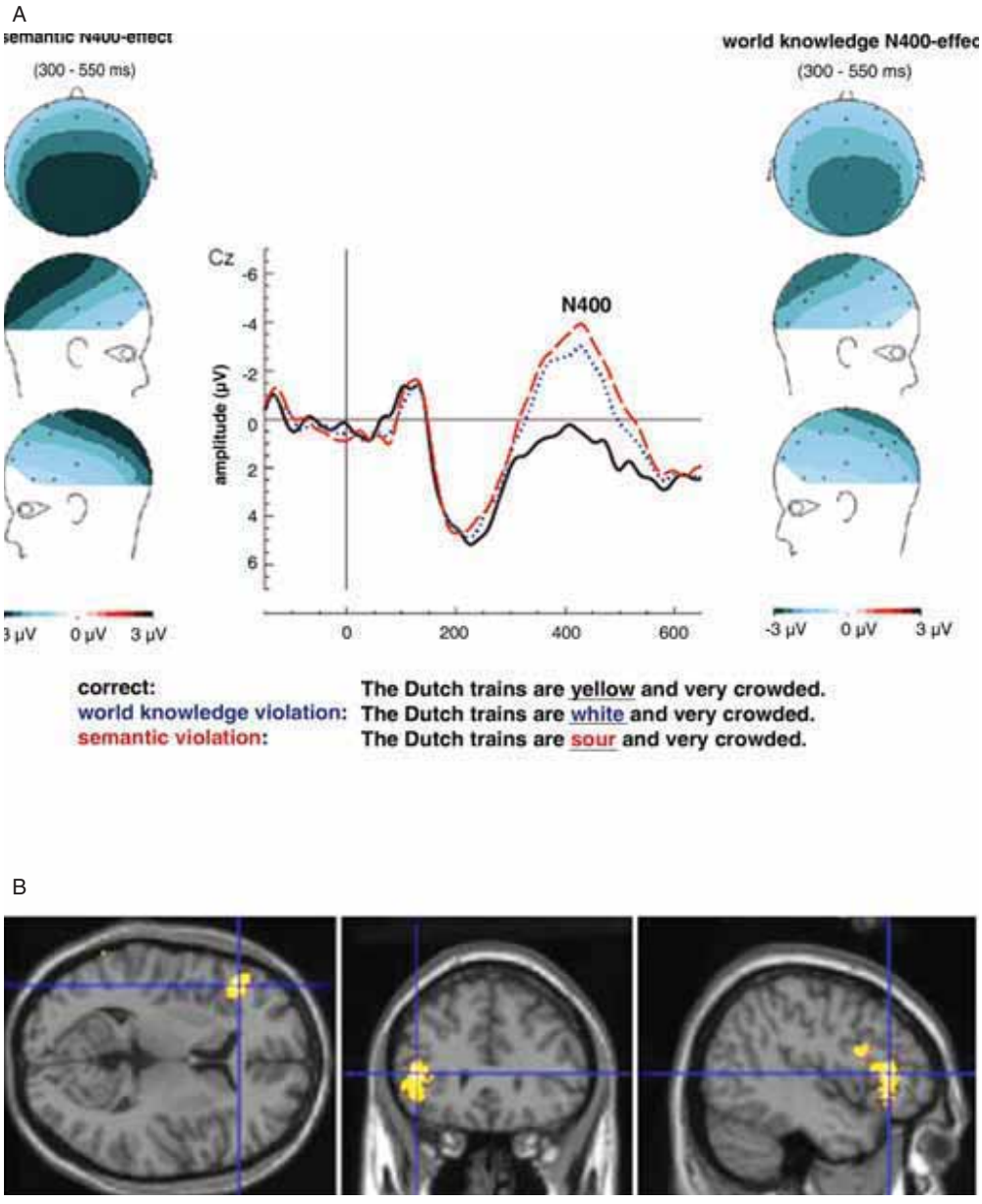


PLATE 71 (*A*) Grand average ERPs for a representative electrode site (Cz) for correct condition (black line), world-knowledge violation (blue dotted line), and semantic violation (red dashed line). ERPs are time locked to the presentation of the critical words (underlined). Spline-interpolated isovoltage maps display the topographic distributions of the mean differences from 300 to 550 ms between semantic violation and control (left), and between world knowledge violation and control (right). (*B*) The common activation for semantic and world-knowledge violations compared to the correct condition, based on the results of a minimum-T-field conjunction analysis. Both violations resulted in a single common activation ($P = 0.043$, corrected) in the left inferior frontal gyrus. The crosshairs indicate the voxel of maximal activation. (Reprinted with permission from Hagoort, Hald, Bastiaansen, & Petersson, 2004.) (See figure 56.2.)

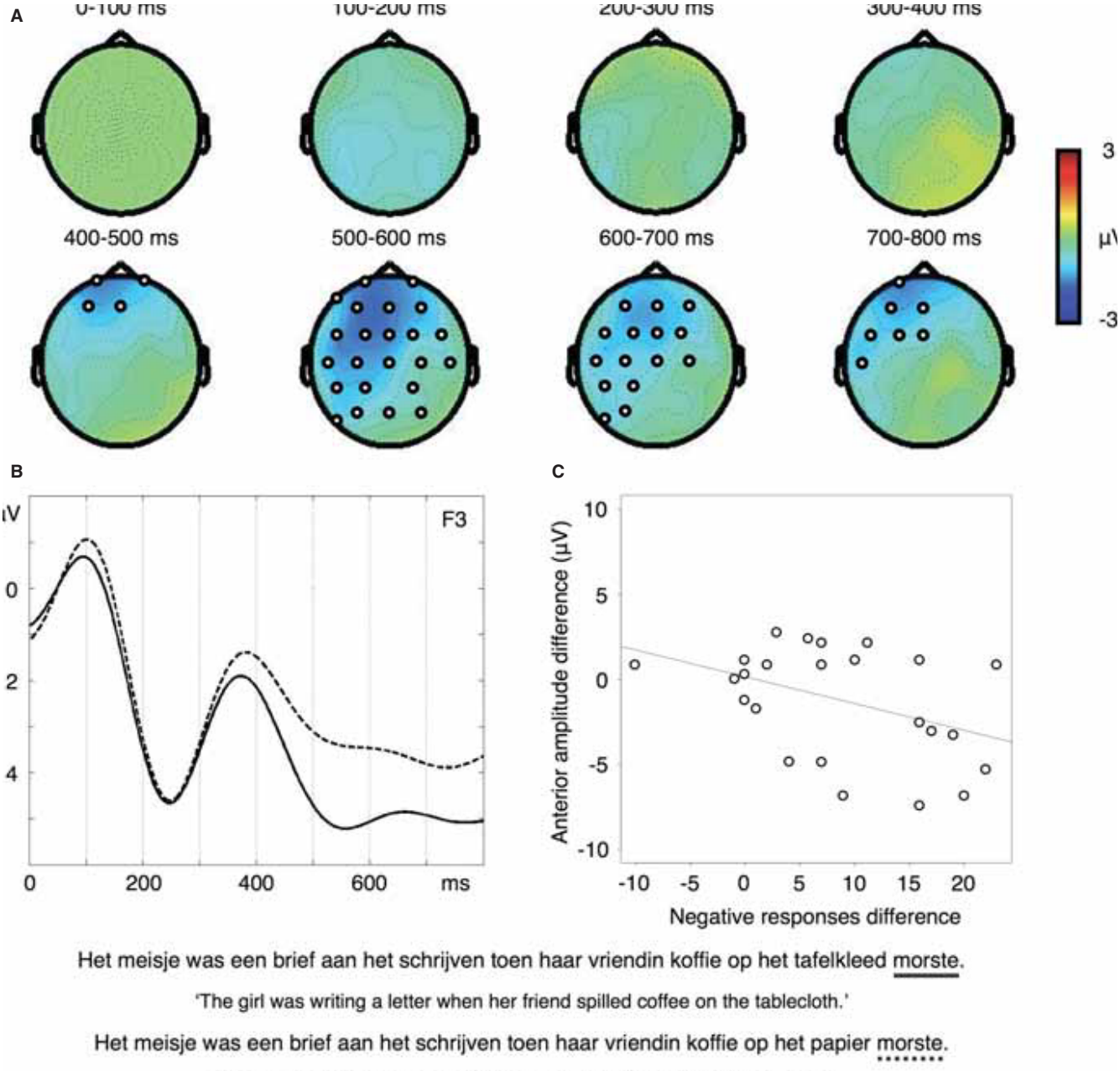


PLATE 72 (*A*) Grand-average topographies displaying the mean amplitude difference between the ERPs evoked by the sentence-final verb when it terminated versus when it did not terminate the accomplishments in the progressive. Circles represent electrodes in a significant ($P < 0.05$) cluster. (*B*) Grand-average ERP waveforms from a representative site (F3) time-locked to the onset (0 ms) of the verb in terminated versus nonterminated accomplishments. Negative values are plotted upward. (*C*) Scatter plot displaying the correlation between the amplitude of the sustained anterior negativity elicited by terminated accomplishments and the frequency of negative responses in a button-press, probe-selection task ($r = -0.415$, $T(22) = -2.140$, $P = 0.043$). The mean difference of negative responses between terminated and nonterminated accomplishments is plotted on the abscissa. The mean amplitude difference at frontopolar and frontal electrodes between terminated and nonterminated accomplishments in the 500–700-ms interval following the onset of the sentence-final verb is plotted on the ordinate. (After Baggio, van Lambalgen, & Hagoort, 2008.) (See figure 56.3.)

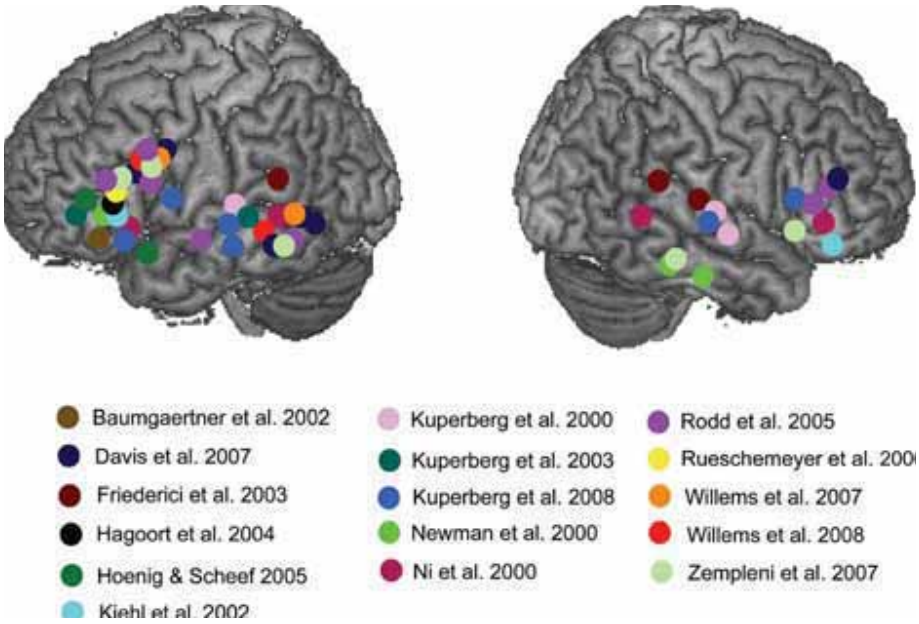


PLATE 73 Overview of local maxima in inferior frontal cortex and in temporal cortex in neuroimaging studies employing sentences with semantic anomalies or semantic ambiguities. The local maxima (in MNI space) of each study were overlaid on a rendering of a brain in MNI space. For local maxima see tables 56.1 and 56.2; for a summary of the results see table 56.3. Rendering was made using MRIcroN. Please note that the local maxima of the Ni and colleagues (2000) and the Kuperberg and colleagues (2003) studies are displayed, but that these are not based on coordinates, since no coordinates were provided. The local maxima are drawn by hand based upon the figures in the respective papers. (See figure 56.6.)

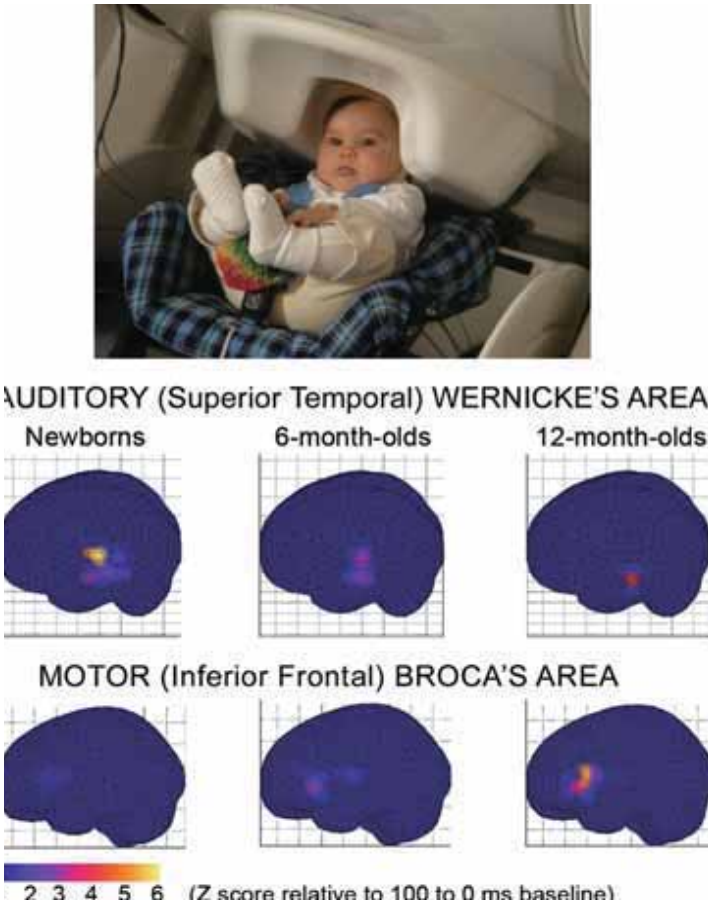


PLATE 74 Neuromagnetic signals were recorded using MEG in newborns, 6-month-old infants, and 12-month-old infants while listening to speech (shown) and nonspeech auditory signals. Brain activation recorded in auditory (*top row*) and motor (*bottom row*) brain regions revealed no activation in the motor speech areas in the newborn in response to auditory syllables. However, activation increased in the motor areas in response to speech (but not nonspeech) in 6- and 12-month-old infants that was temporally synchronized between the auditory and motor brain regions. (From Imada et al., 2006.) (See figure 57.7.)

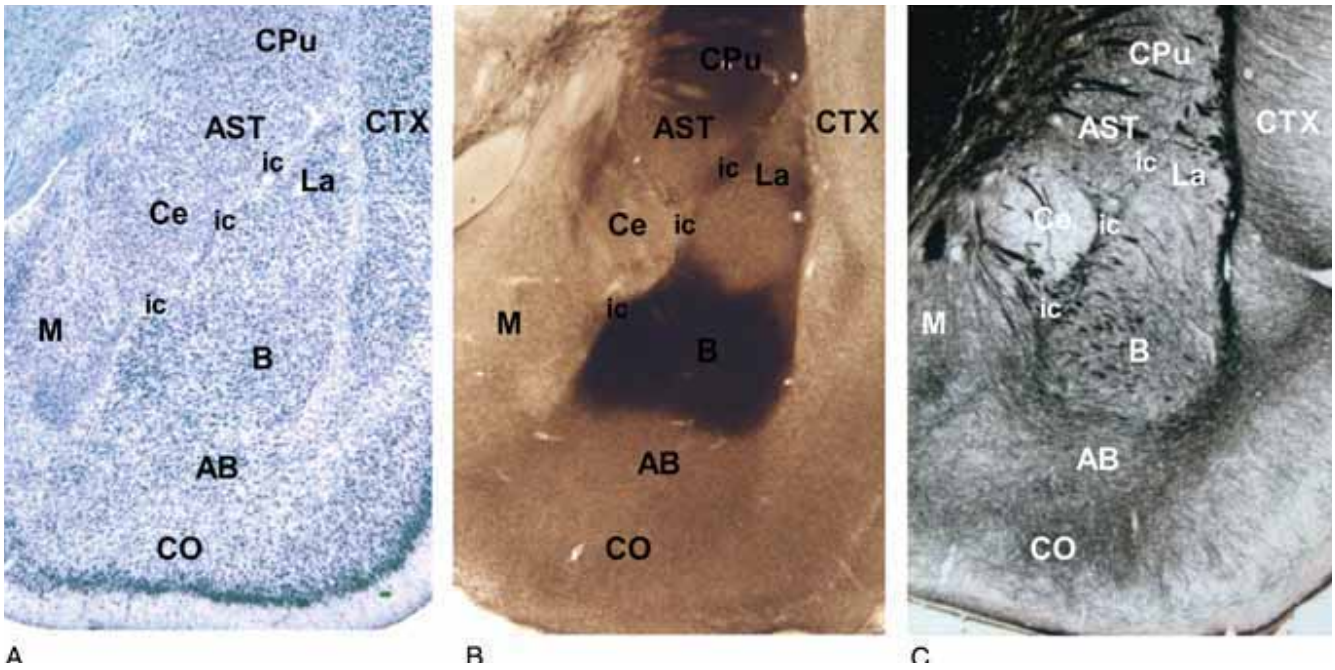


PLATE 75 The rat amygdala. The amygdala of mammals, including humans, consists of at least 12 distinct nuclei. Different staining methods show some of the major nuclei from different perspectives. (*A*) Nissl cell body stain. (*B*) Acetylcholinesterase stain. (*C*) Silver fiber stain. Abbreviations: Amygdala areas: AB, accessory basal; B, basal nucleus; Ce, central nucleus; CO, cortical nucleus; ic, intercalated cells; La: lateral nucleus; M, medial nucleus. Nonamygdala areas: AST, amygdalo-striatal transition area: CPu, caudate putamen; CTX, cortex. (See figure 61.2.)

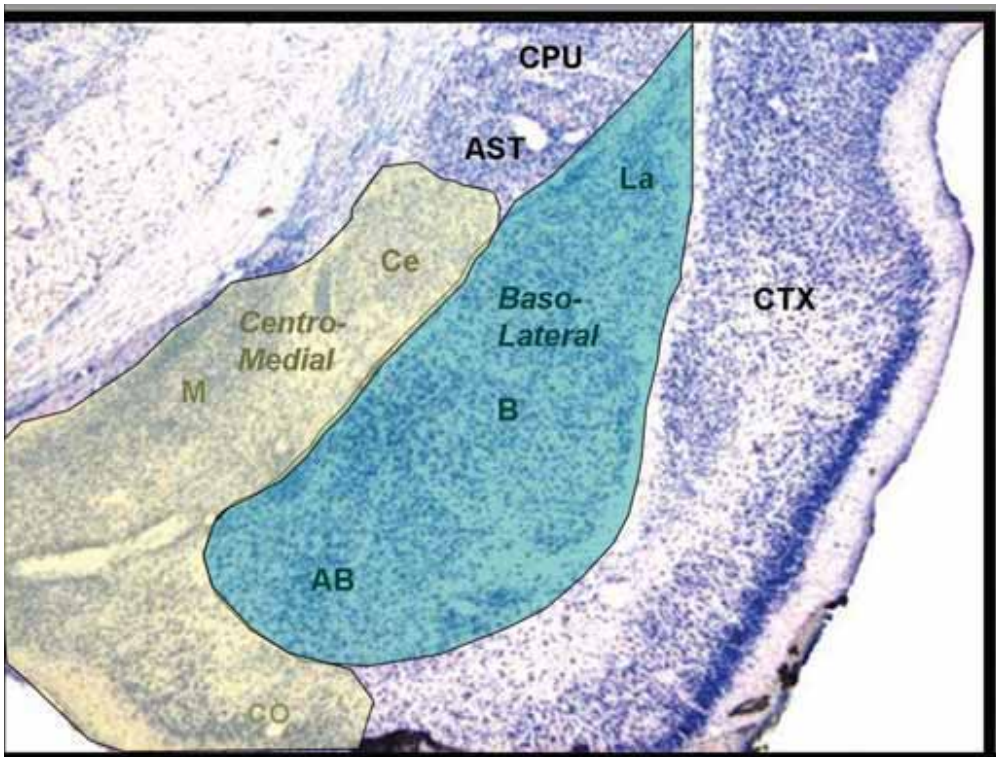


PLATE 76 Groupings of amygdala nuclei. The various nuclei of the amygdala are often partitioned into an evolutionarily old division (the centromedial or corticomedial region) and an evolutionarily newer division (the basolateral region or basolateral complex). While these divisions have some value in understanding the phylogenetic and ontogenetic origins of the amygdala, they do not represent meaningful function divisions, since functions are mediated by cells within much more localized regions, especially subnuclei and even subdivisions of subnuclei. Abbreviations: AB, accessory basal; AST, amygdalo-striatal transition area; B: basal nucleus; Ce, central nucleus; CPu, caudate putamen; CTX, cortex; La, lateral nucleus; M, medial nucleus. (See figure 61.3.)

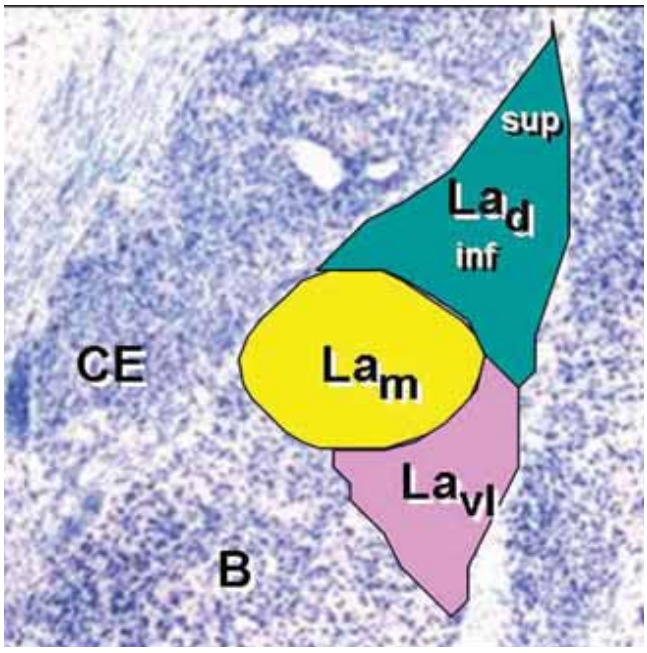


PLATE 77 Subdivisions of the lateral nucleus of the amygdala. The lateral nucleus of the amygdala has three major subdivision: dorsal (LA_d), ventrolateral (LA_{vl}), and medial (LA_m). Each of these has additional partitions. The dorsal subnucleus, for example, contains a superior (sup) and inferior (inf) region. Cells in the superior region have been implicated in the acquisition of fear conditioning, and cells have been implicated in the inferior region in long-term memory storage (see text). Abbreviations: B, basal nucleus; CE, central nucleus. (See figure 61.4.)

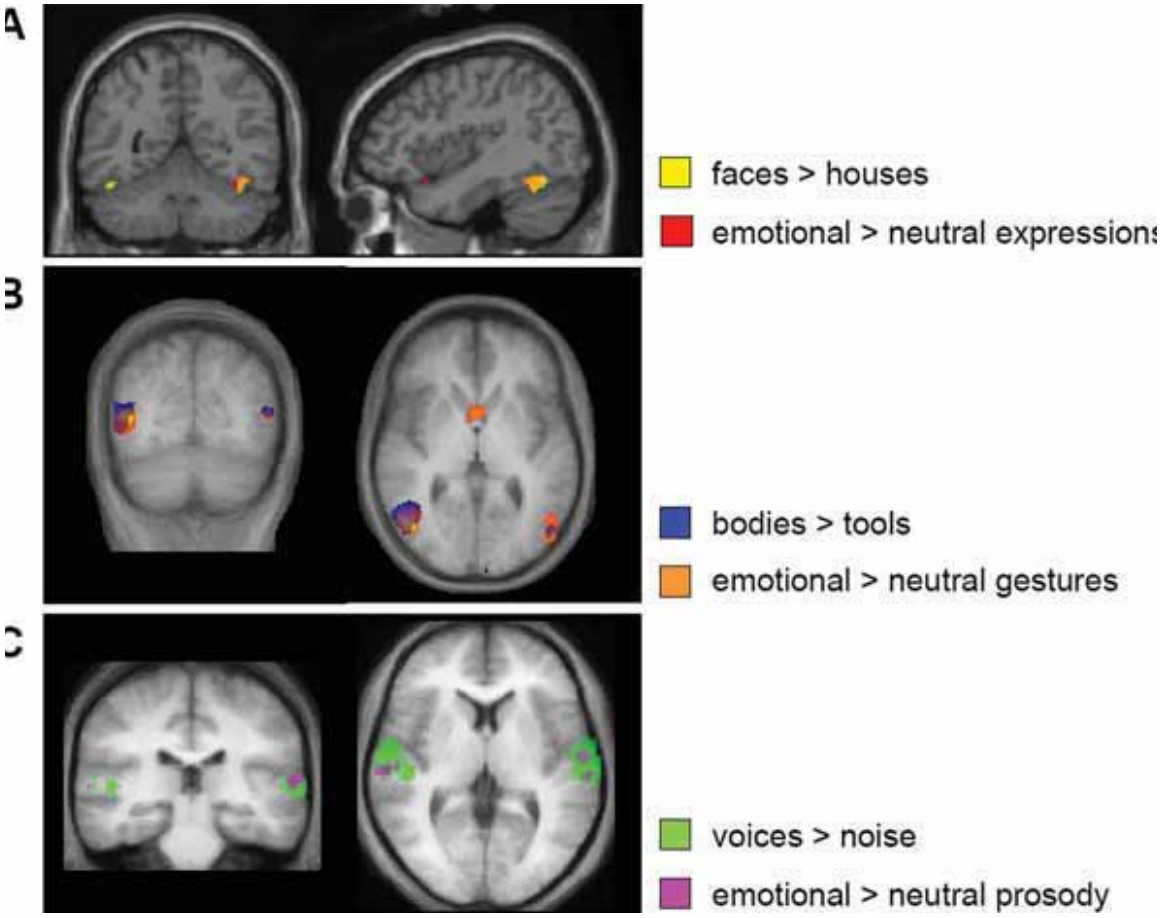


PLATE 78 Emotional enhancement of neural responses in fMRI. (*A*) Faces with a fearful relative to neutral expression produce increased activation in fusiform cortex, overlapping with the fusiform area selectively activated by faces as compared with houses. (From Vuilleumier et al., 2001.) (*B*) Bodies with dynamic gestures expressing various emotions (fear, anger, happiness, or disgust) produce increased activation in lateral occipital area cortex, overlapping with the extrastriate area selectively activated by bodies as compared with tools. (From Peelen, Atkinson, Andersson, & Vuilleumier, 2007.) (*C*) Voices with angry prosody produce increased activation in temporal cortex, overlapping with an area in superior temporal gryus selectively activated by human voices as compared with noises with similar acoustic energy. (From Grandjean et al., 2005.) (See figure 62.1.)

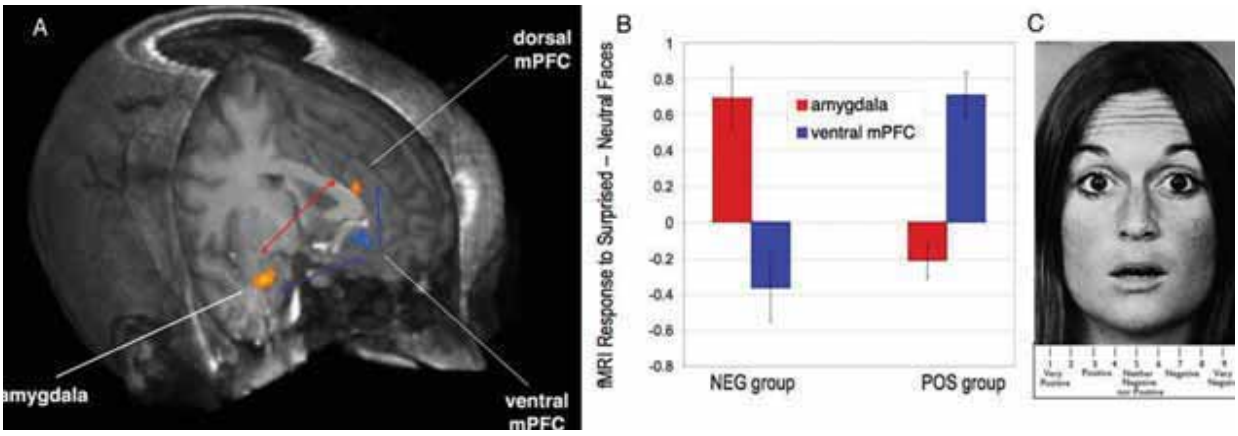


PLATE 79 (*A*) A three-dimensional depiction of the correlational results of H. Kim and colleagues (2003). Amygdala and dorsal mPFC loci that showed a positive correlation with valence ratings of surprise (colored in orange) are also positively correlated with one another (red arrow, $r = +.66$). The ventral mPFC locus that showed a negative correlation with valence ratings of surprise (colored in blue) is also negatively correlated with amygdala (blue arrow; $r = -.69$) and the dorsal mPFC (blue arrow; $r = -.62$). (*B*) Bar graph focusing on the inverse relationship between the amygdala and ventral mPFC in subjects who interepted the surprised faces either positively (POS) or negatively (NEG). (*C*) An example of the surprised faces and valence scale used to rate them.(See figure 63.2.)

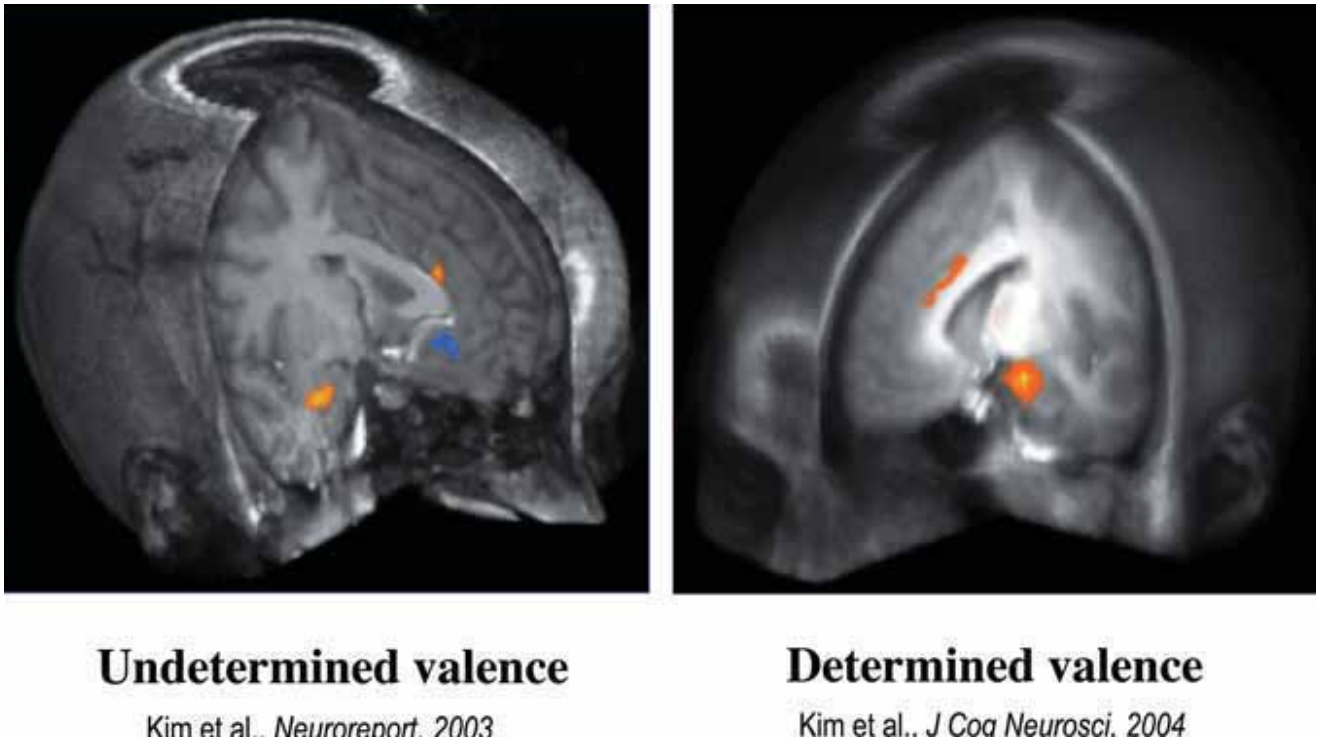


PLATE 80 Comparison of results of H. Kim and colleagues (2003) with those of H. Kim and colleagues (2004). (See figure 63.4.)

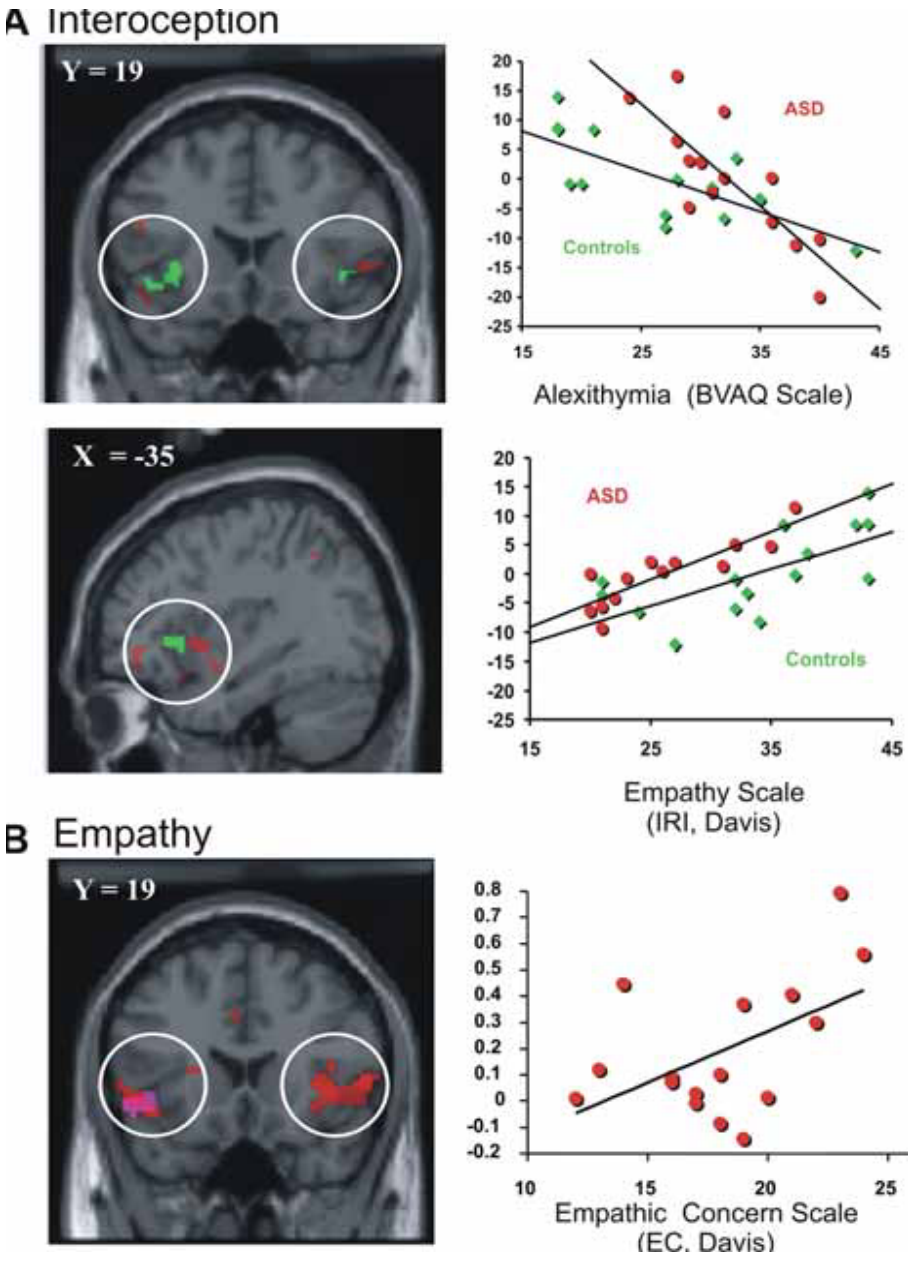


PLATE 81 (*A*) Bilateral AI activation during interoception about one's own feelings correlates with the degree of alexithymia (measured with the Bermond-Vorst Alexithymia Questionnaire by Vorst and Bermond, 2001; BVAQ-B) and empathy (measured with the Interpersonal Reactivity Index by Davis, 1980; IRI) in healthy controls and in subjects with Asperger syndrome (adapted from Silani et al., 2008). (*B*) Overlapping brain activity in bilateral AI in females experiencing pain themselves and empathizing with their partner (red; adapted from Singer et al., 2004) or an unfamiliar but likable person (pink; adapted from Singer et al., 2006) experiencing pain. Individual differences in self-reported empathy (measured with the IRI) covary with activation strengths in AI during empathizing. (From Singer et al., 2004.) (See figure 67.1.)

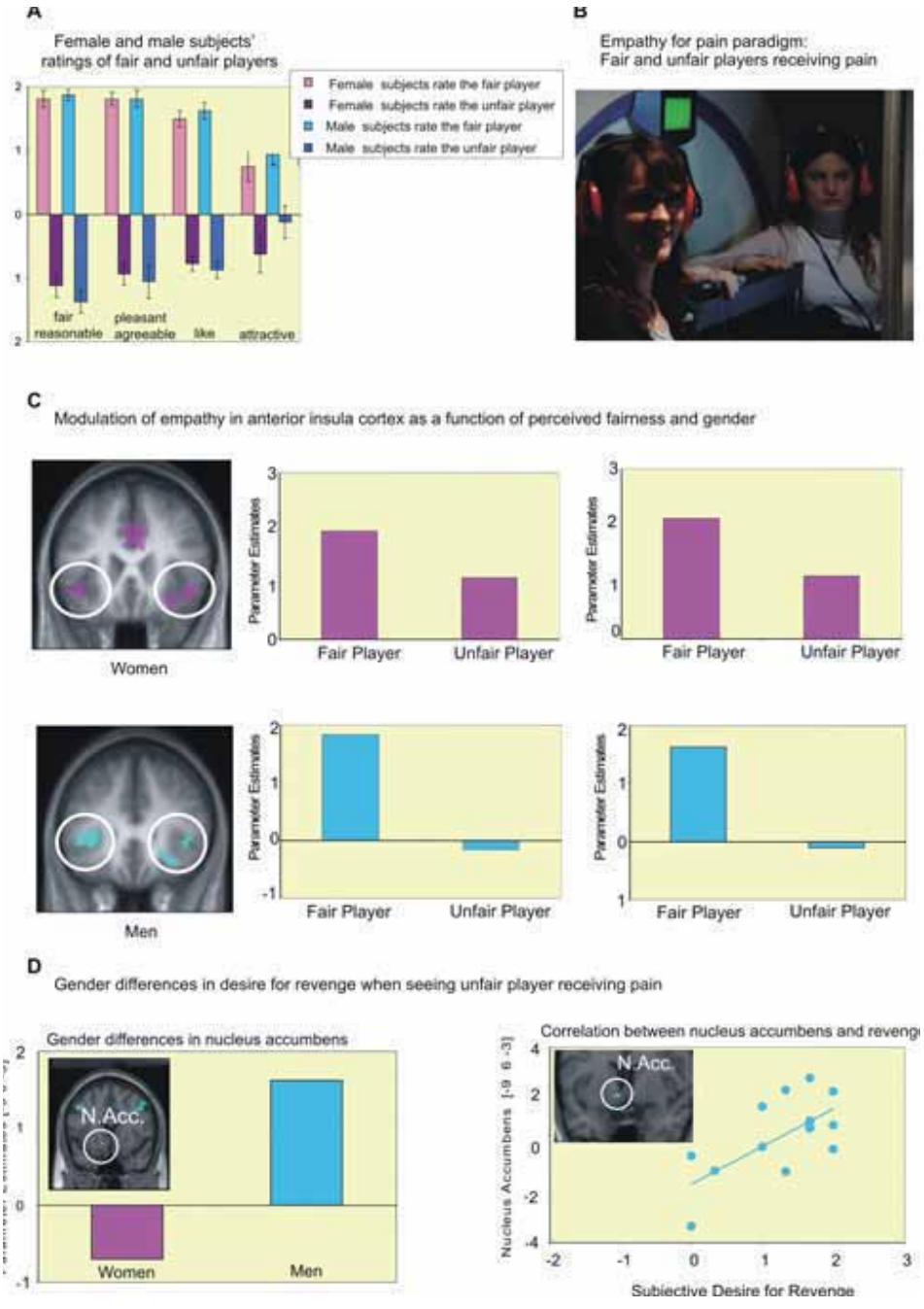


PLATE 82 Modulation of empathic brain responses by perceived fairness (adapted from Singer et al., 2006). (*A*) Female (pink) and male (bule) subject's postscan ratings of perceived fairness, agreeableness, likeability, and attractiveness of the two confederates, who had always played fairly and unfairly, respectively, in a preceding monetary economic trust game. (*B*) Setup of the empathy-for-pain paradigm with the subject lying in the scanner and one fair and one unfair player (both confederates) sitting on either side of the scanner. Electrodes, which were attached to one of their hands, delivered painful or nonpainful stimulation as previously indicated by flashes on a screen in front of them. (*C*) Empathic brain activation in bilateral frontoinsular cortex when males (blue) and females (pink) perceived either the fair or the unfair player suffering pain. Although both men and women showed empathic brain activity in frontoinsular cortex when they perceived a fair player in pain, only women did so when they perceived an unfair player in pain. (*D*) Enhanced activation in nucleus accumbens in men, but not in women, when they perceived unfair as compared to fair players suffering pain. The strength of this activation correlated positively with men's , but not women's, degree of subjectively expressed desire for revenge.(See figure 67.2.)

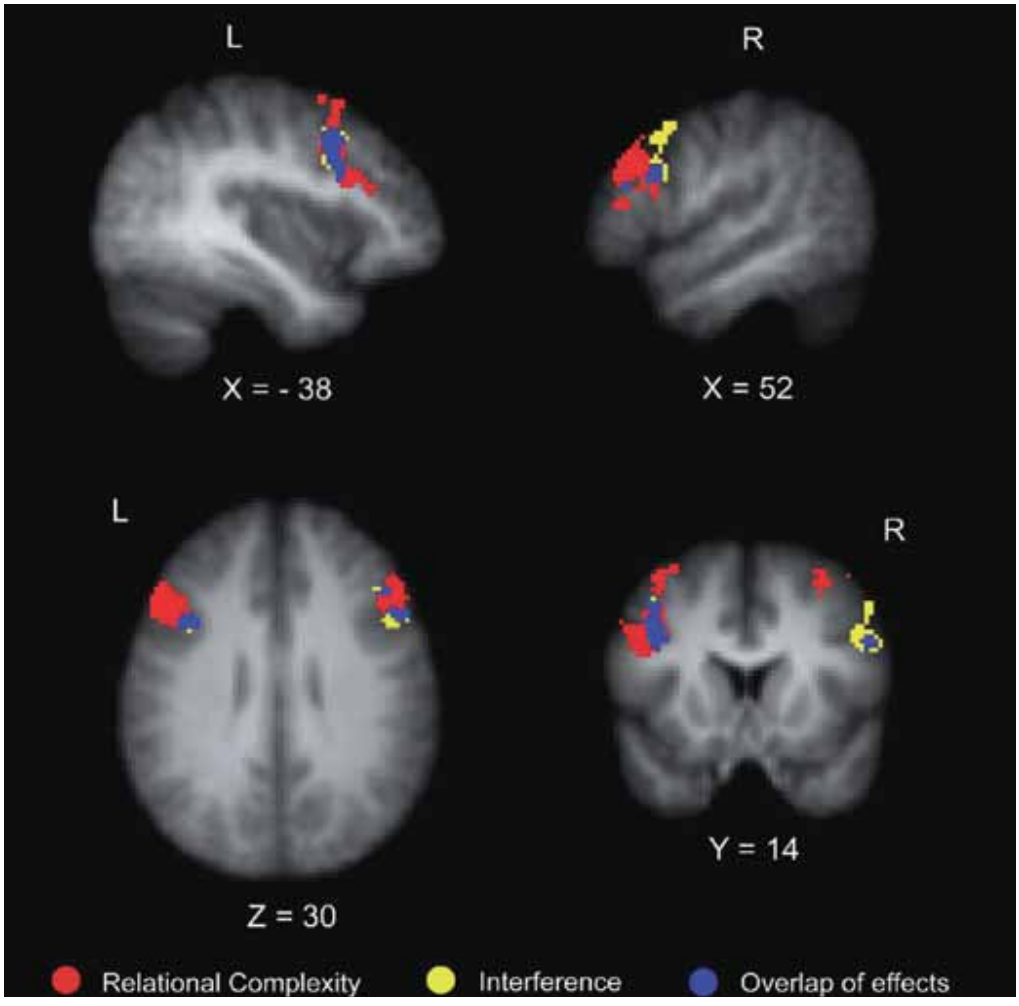


Plate 83 Neuroimaging results from Cho and colleagues (2009). Regions showing the main effects of relational complexity (shown in red), interference (shown in yellow; small volume corrected, uncorrected cluster-forming threshold $T > 2.3$, corrected cluster extent significance threshold, $p < .05$), and regions where main effects overlapped (blue) within an a priori defined anatomical ROI mask of the bilateral MFG and IFG pars opercularis and pars triangularis. R, right; L, left. Coordinates are in MNI space (mm). (See figure 69.7.)

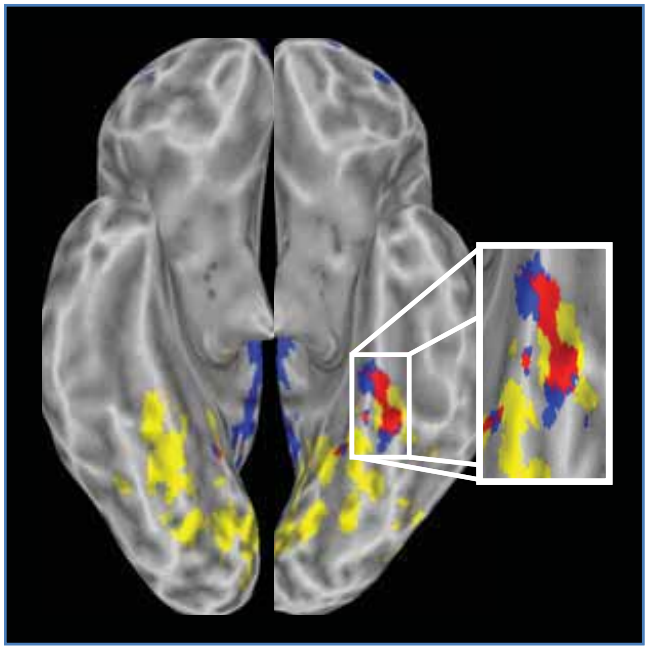


Plate 84 Overlap between the neural circuitry for perceiving and knowing about color. Shown is an inflated map of the ventral surface of the brain. Regions shown in yellow were more active when subjects performed a difficult color-perception task, relative to performing that same task with gray-scale stimuli. Regions in blue were more active when answering written questions about object color, relative to answering questions about object motor and motion properties. Red shows region of overlap in the left fusiform gyrus for the color-perception and color-knowledge tasks. (Adapted from Simmons, Ramjee, McRae, Martin, & Barsalou, 2007.) (See figure 71.2.)

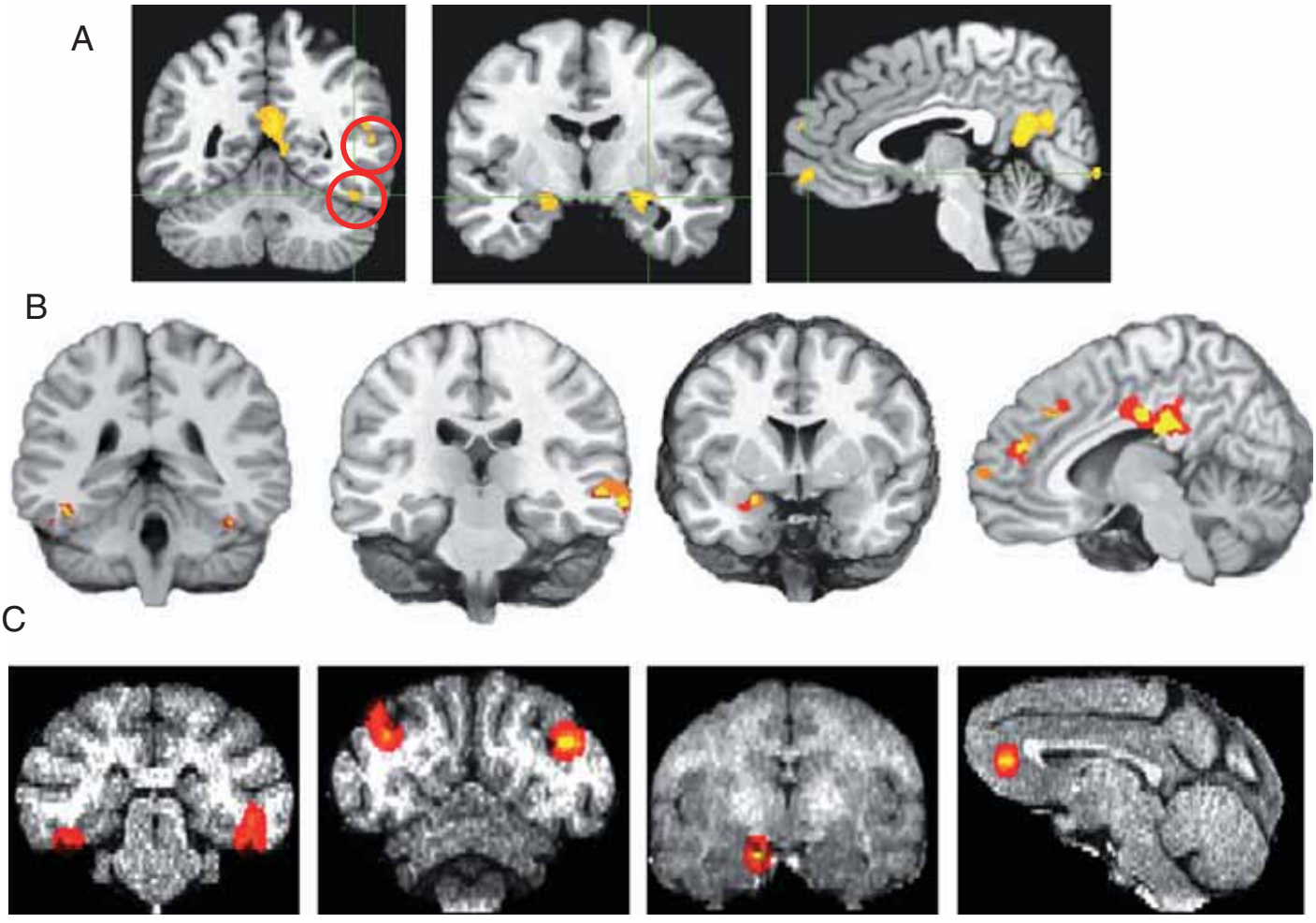


PLATE 85 Correspondence across tasks and species in the location of the neural circuitry for perceiving and knowing about animate entities. (*A*) Regions shown in yellow were more active when subjects viewed photographs of faces relative to viewing photographs of common tools. Going from left to right, the first image shows a coronal slice through posterior cortex indicating the location of activity in the lateral portion of the right fusiform gyrus (lower red circle) and in the right pSTS (upper red circle). The next coronal image depicts bilateral activity in the amygdalae. The third image shows a sagittal section revealing activity in the medial prefrontal cortex and in the posterior cingulate/precuneus. (Unpublished data from our laboratory.) (*B*) Brain slices depicting conjunction of regions more active when subjects perceived simple shapes in motion as animate, relative to when they were judged to be inanimate, and when they imagined these stimuli as animate versus inanimate. Going from left to right, the first image is a coronal slice showing bilateral activity in the lateral fusiform gyrus. The next coronal slice shows the location of activity in the STS, the third depicts activity in the left amygdala, and the last shows activations located in the medial prefrontal and posterior cingulate cortices. (Adapted from Wheatley, Milleville, & Martin, 2007.) (*C*) Activity in the macaque brain when listening to species-specific calls. Shown are PET scans obtained from a single animal. Going from left to right, the first image shows a coronal slice through ventral regions TEO/TE, the next coronal slice shows activity in the STS, the third slice shows activation in the amygdala, and the fourth slice shows an activation located in Area 32 on the medial surface of the brain. (Adapted from Gil-da-Costa et al., 2004.) (See figure 71.3.)

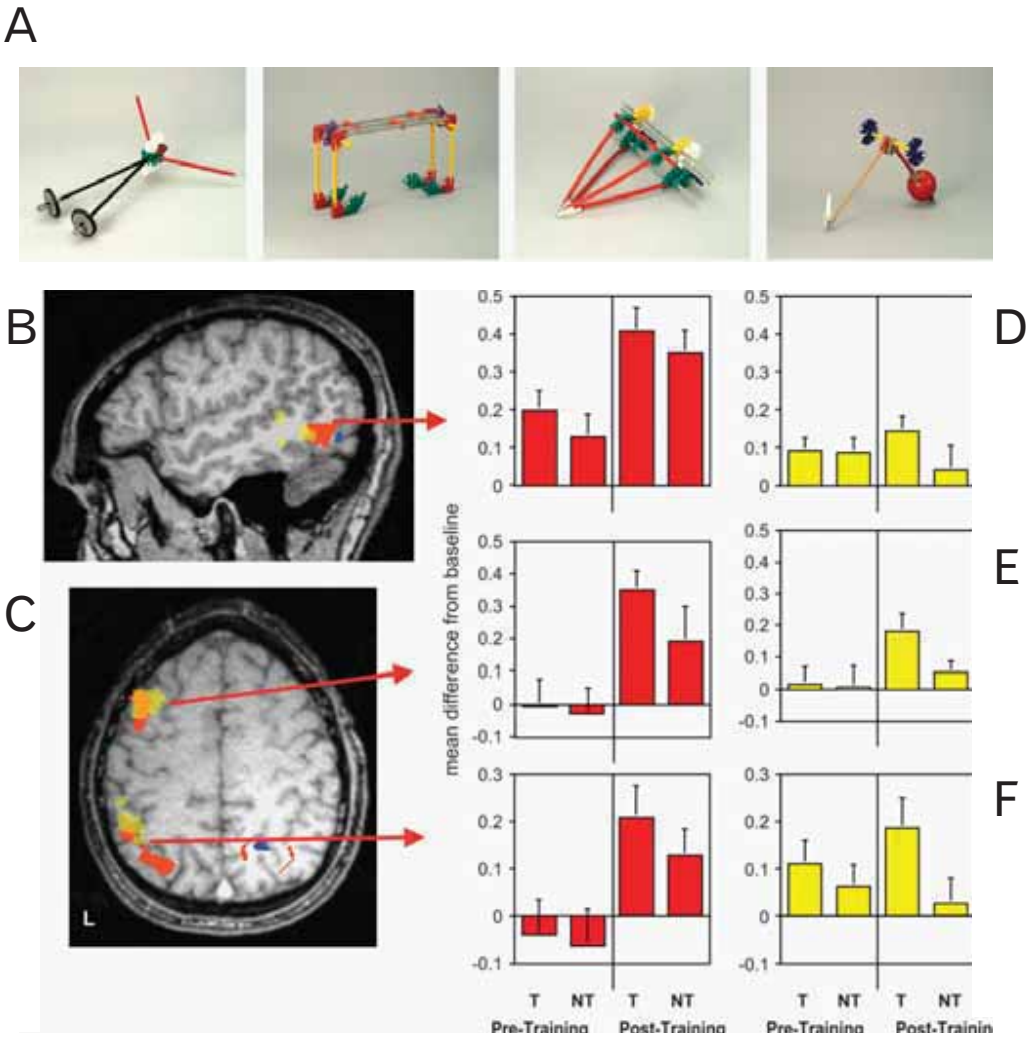


PLATE 86 (*A*) Examples of novel objects designed to perform specific toollike functions. (*B*) Sagittal section showing the location of learning-related activity in the left middle temporal gyrus. Regions in red were more active after training than before training. Regions in yellow, which overlap with regions in red, were more active for trained (*T*) objects than for not-trained (*NT*) objects. (*C*) Axial section showing the location of learning-related activity in the left premotor/prefrontal cortex and intraparietal cortices. (*D*, *E*, *F*) Histograms showing the difference between novel-object-matching and scrambled-image-matching baseline task in the middle temporal gyrus, left premotor, and intraparietal regions, respectively. Red bars represent brain regions that showed increased activity for object matching after but not prior to training; yellow bars represent regions that demonstrated greater activity for trained objects than not-trained objects after but not prior to training. (Adapted from Weisberg, van Turennout, & Martin, 2007.) (See figure 71.4.)

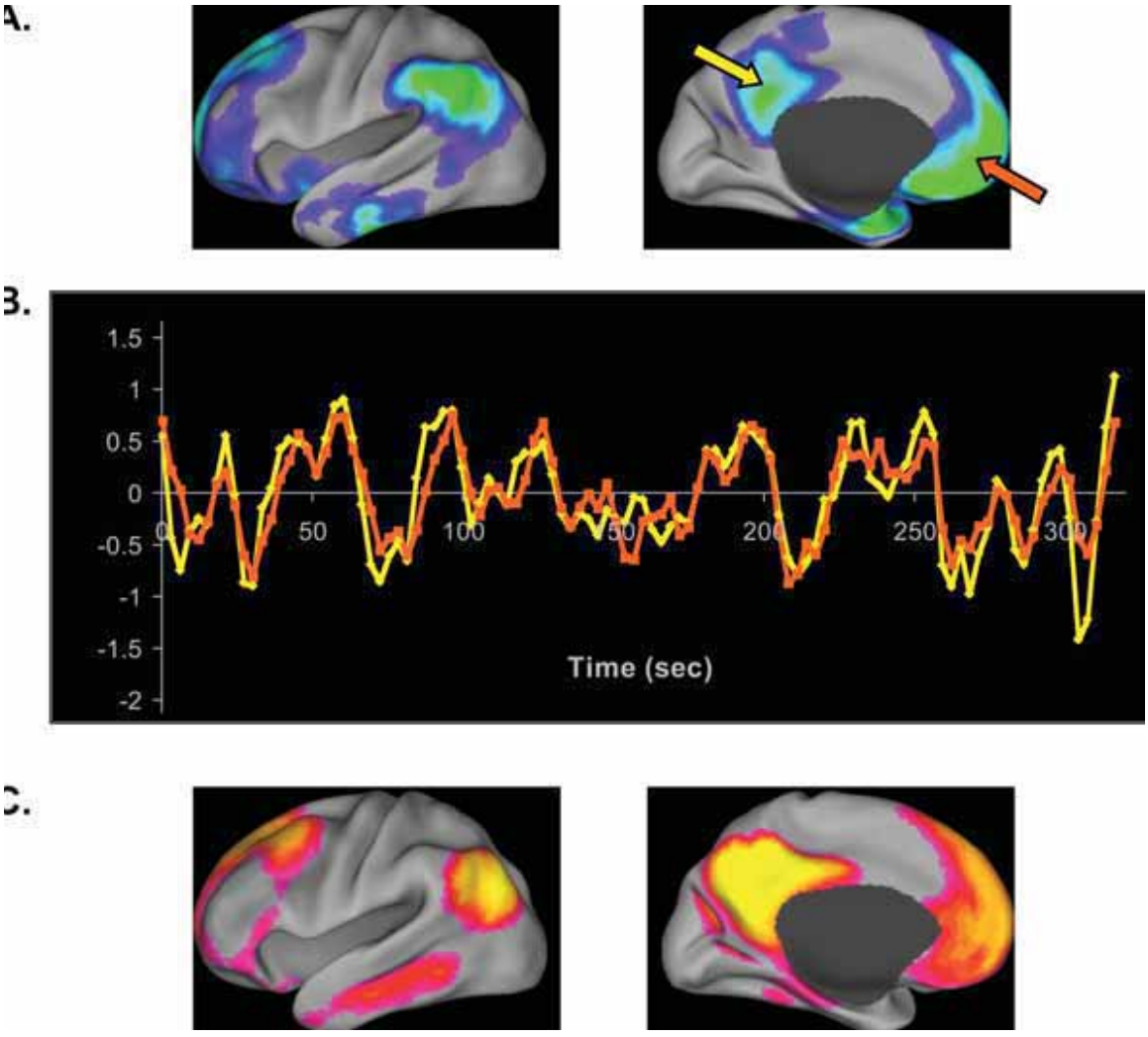
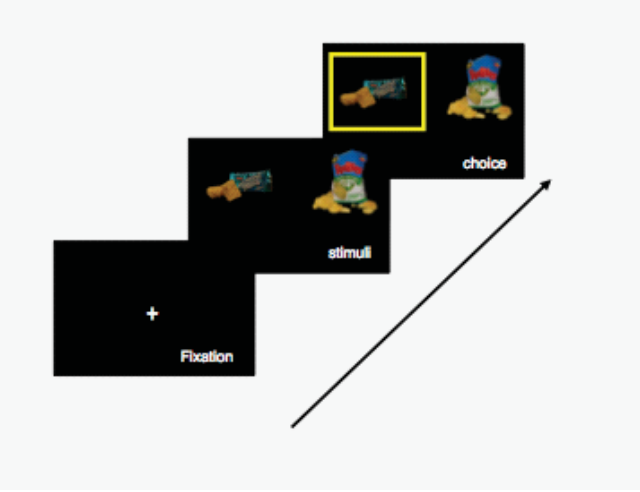


PLATE 87 Performance of a wide variety of tasks has called attention to a group of brain areas (*A*) that decrease their activity during task performance (data adapted from G. Shulman et al., 1997). These areas are often referred to as the brain's default mode network after our initial work on them (Raichle et al., 2001). If one records the spontaneous fMRI BOLD signal activity in these areas in the resting state (arrows, *A*) what emerges is a remarkable similarity in the behavior of the signals between areas (*B*), a phenomenon originally described by Biswal and colleagues (1995) in the somatomotor cortex and later in the default mode network by Greicius and colleagues (2003). Using these fluctuations to analyze the network as a whole (M. Fox et al., 2005; Vincent et al., 2006) reveals a level of functional organization (*C*) that parallels that seen in the task-related activity decreases. These data provide a dramatic demonstration that the ongoing organization of the human brain likely provides a critical context for all human behaviors. (These data were adapted from our earlier published work: M. Fox et al., 2005; Gusnard & Raichle, 2001; Raichle et al., 2001; G. Shulman et al., 1997.) (See figure 73.1.)

A



B

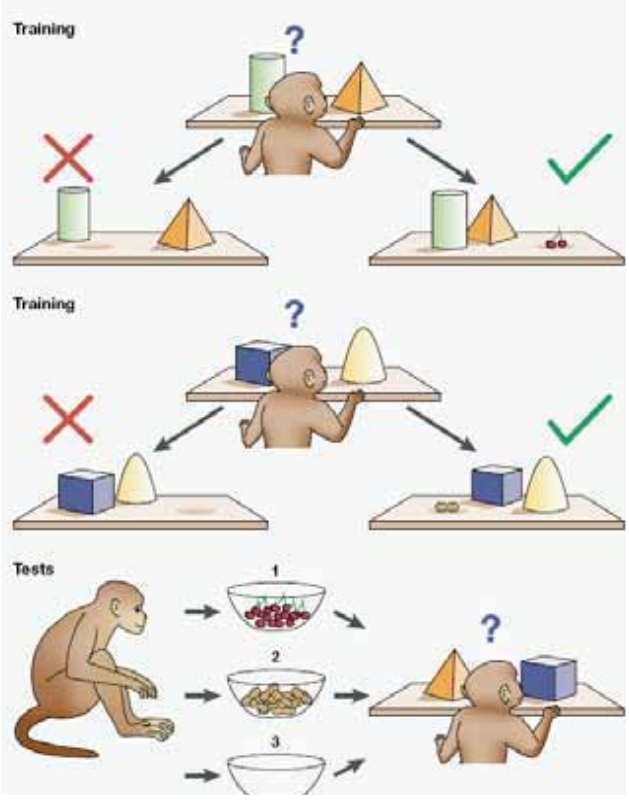


PLATE 88 Examples of simple binary stimulus choice tasks. (*A*) Binary food choice from Karjbich, Armel, and Rangel (under review). (*B*) Devaluation choice task from Izquierdo, Suda, and Murray (2004). (With permission from Baxter & Murray, 2002.) (See figure 74.1.)

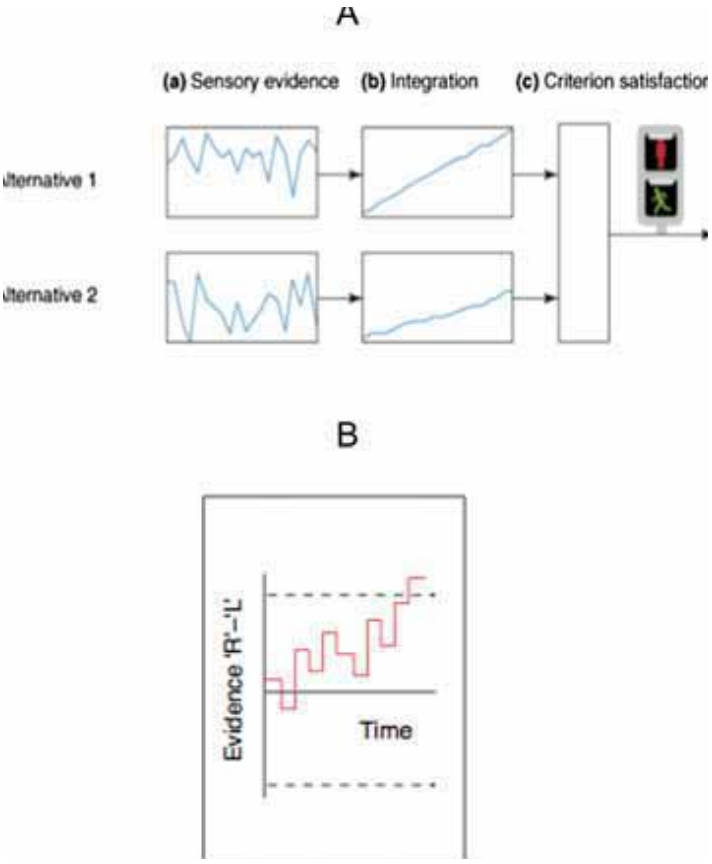


PLATE 89 Models of the value comparison process. (*A*) Illustration of the main components of the race-to-barrier models. (Adapted with permission from Bogacz, 2007.) (*B*) A typical run of the random walk model. The step function represents the accumulated relative value of the "right" target. The process starts at a middle point and stops the first time this variable crosses one of the thresholds (depicted by the bracketed horizontal lines). "Right" is chosen when it crosses the upper threshold; "left" is chosen when it crosses the lower one. Time advances in discrete steps. The size of every step is given by a Gaussian distribution with a mean that is proportional to the true direction of motion. This noise is meant to capture the variability in the valuation processes. (See figure 74.2.)

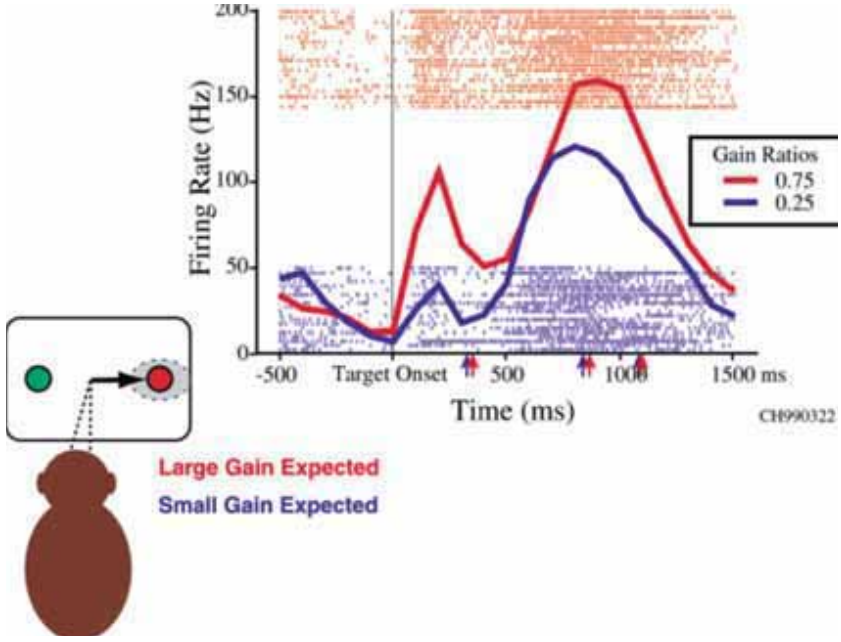


PLATE 90 Same movement, different values. (From Glimcher, 2003.) (See figure 75.1.)

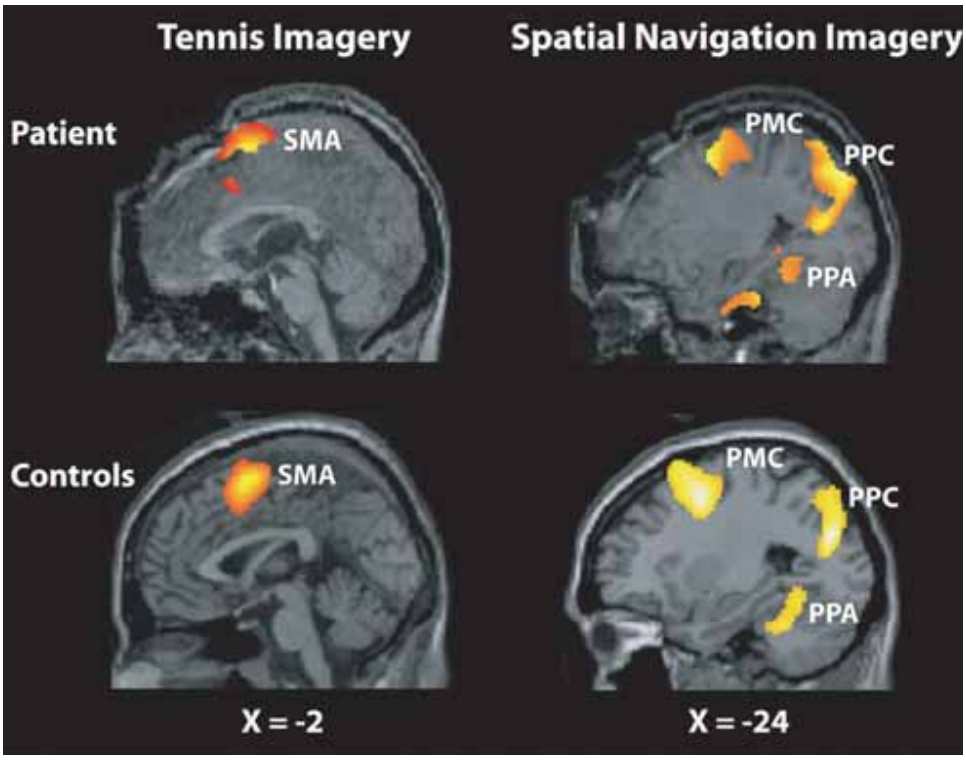


PLATE 91 Command following in vegetative state (Owen et al., 2006). A 23-year-old woman with clinical exam consistent with VS, five months after severe traumatic brain injury with only brief periods of visual fixation, was asked to imagine playing tennis or walking throughout her own house. The regionally selective brain activation patterns obtained from functional magnetic resonance imaging measurements for each condition were identical to those of normal controls. (Reproduced with permission.) (See figure 78.2.)

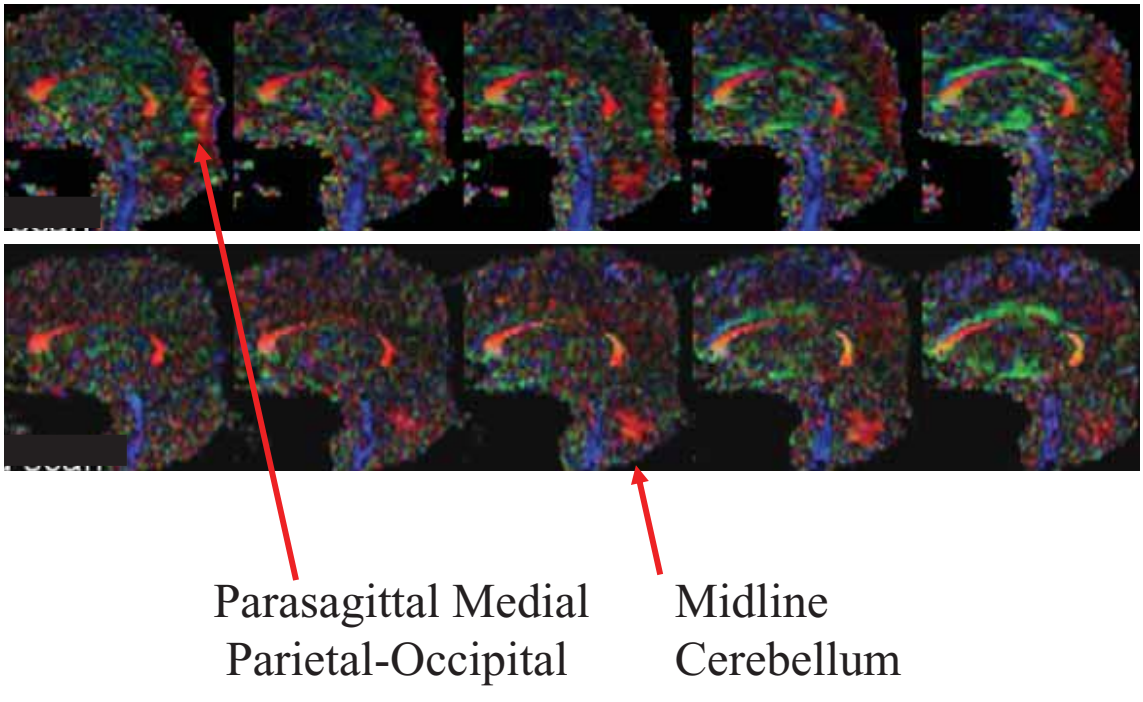
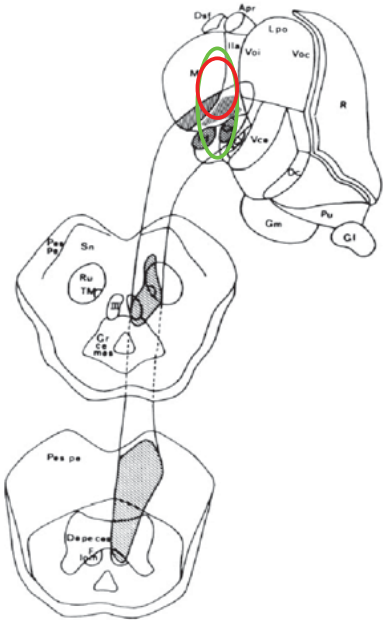
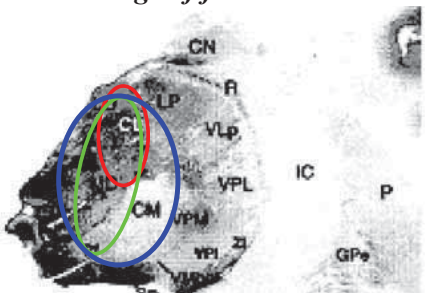


PLATE 92 Diffusion tensor imaging studies of patient with late recovery (19 years) from MCS (Voss et al., 2006). Fractional anisotropy maps showing fiber tracks: red, fibers with left-right directionality; blue, fibers with up-down directionality; green, fibers with anterior-posterior directionality. Top images show volume loss of the corpus callosum throughout the medial component and regions in parieto-occipital white matter with prominent left-right directionality. Bottom row images show fractional anisotropy maps obtained 18 months later that demonstrate reduction of left-right direction in parieto-occipital regions with increased anisotropy noted in the midline cerebellum. (See figure 78.3.)

A. Anatomical localization of focal subcortical injuries producing acute coma, persistent vegetative state and minimally conscious state



B. Regional neuronal cell loss in central thalamus across range of functional outcomes



•Moderately disabled

•Severely disabled

•Vegetative

PLATE 93 Contributions of the central thalamus to disorders of consciousness. (*A*) Focal injury patterns in the central thalamus associated with coma, vegetative state, and minimally conscious state. (*B*) Regional neuronal cell loss in central thalamus across the range of function outcomes. Moderately disabled (red): neuronal loss in median dorsalis, rostral central medial, central lateral and paracentral nuclei. Severely disabled (green): includes moderately disabled regions plus neuronal loss from median dorsalis, caudal central medial, parafasicular nucleus. Permanent vegetative state (blue): all of above and centromedian nucleus. (Figure elements adapted from Castaigne et al., 1981, and Münkle, Waldvogel, & Faull, 2000.) (See figure 78.5.)

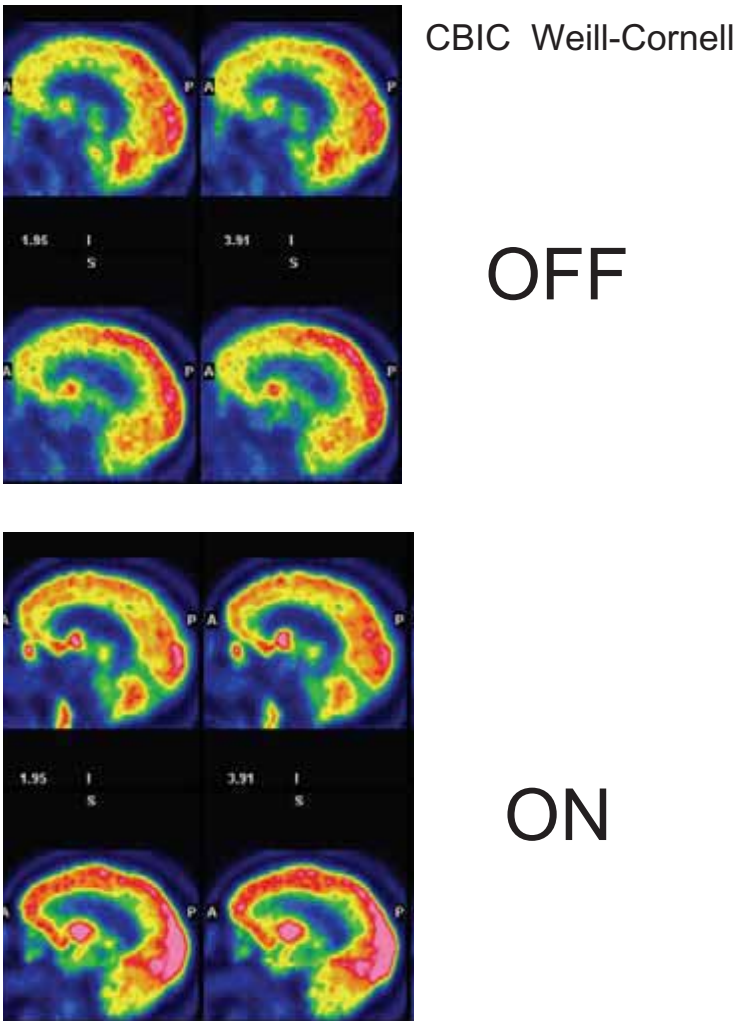


PLATE 94 Changes in cerebral metabolism associated with zolpidem administration in severe brain injury. Two sets of parasagittal images of fluorodeoxyglucose positron emission tomography studies obtained in an awake, severely brain-injured patient are shown obtained one day apart. The OFF images show the resting metabolic profile in an awake state and demonstrate qualitative downregulation of metabolism in the anterior forebrain (frontal, prefrontal cortex), basal ganglia, and thalamus. The ON images show the resting metabolic profile in an awake state 45 minutes after administration of the drug zolpidem. A marked qualitative increase in metabolism is observed in the anterior forebrain, basal ganglia, and thalamus. In addition, overall metabolic activity across cerebral structures is increased. (See figure 78.7.)

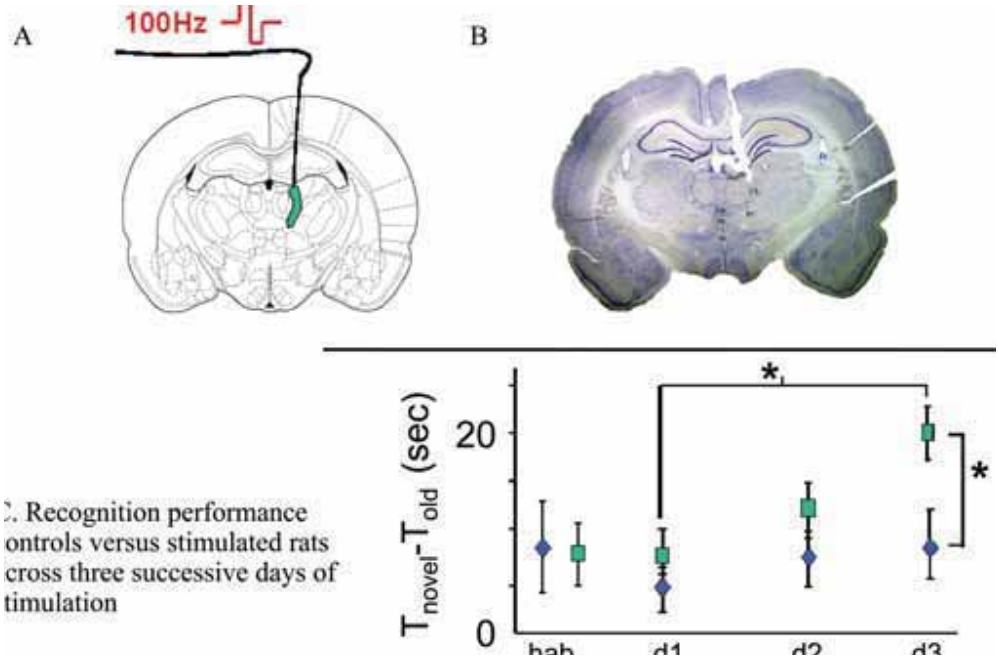


PLATE 95 Facilitation of recognition memory with continuous electrical stimulation of the rat central thalamus (Shirvalkar, Seth, Schiff, & Herrera, 2006). (*A*) Electrode placement in the central thalamus. (*B*) Representative gross histological section. (*C*) Comparison of sham-stimulated controls and stimulated rat cohorts on object recognition task across three successive days of stimulation. Stimulated rats show increased performance and accumulation of effects. (See figure 78.9.)

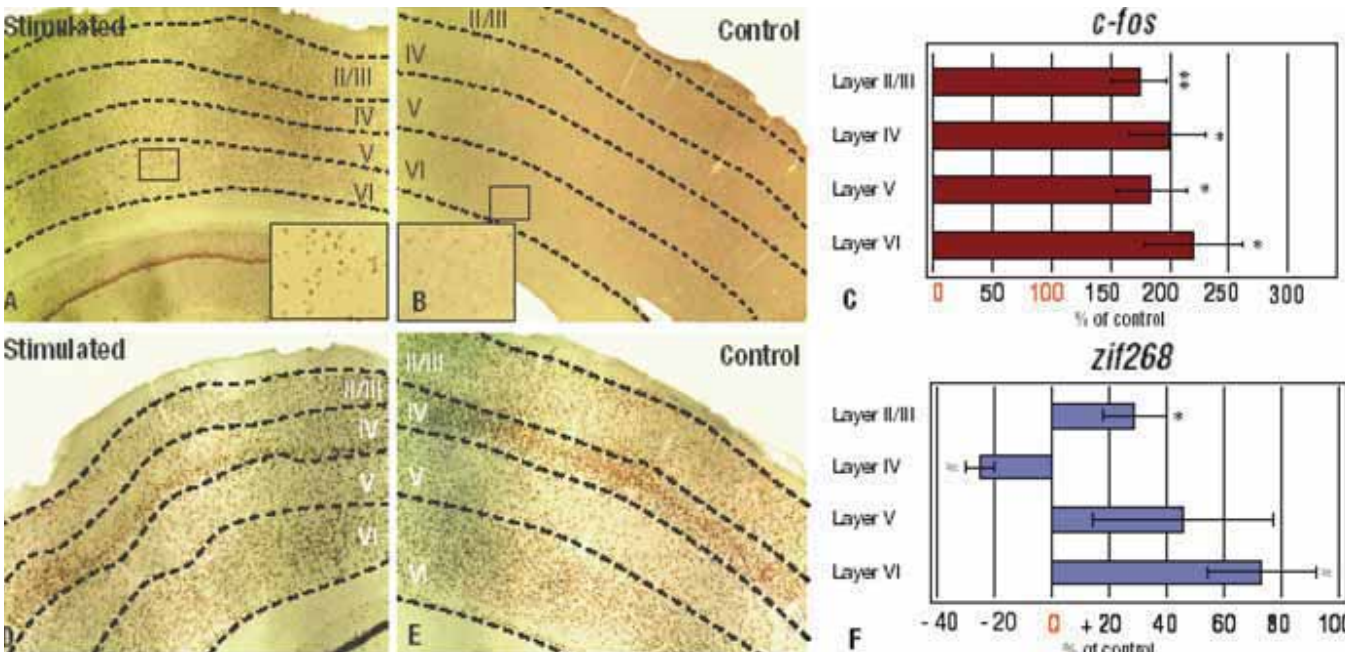


PLATE 96 Gene expression changes in the motor cortex following central thalamic electrical stimulation. Gene expression changes associated with 30 minutes of electrical stimulation of the central thalamus in two immediate early genes, *c-fos* and *zif268*, are shown (Shirvalkar et al., 2006). A broad increase of *c-fos* across cortical lamina is seen, consistent with increased synaptic activity. A laminar specific pattern of changes in *zif268*, a memory-related gene, may link to accumulation effects seen in figure 78.9 and possibly to carryover effects observed in human-subject DBS study shown in figure 78.8 (see text and Shirvalkar et al., 2006, for further details). (See figure 78.10.)

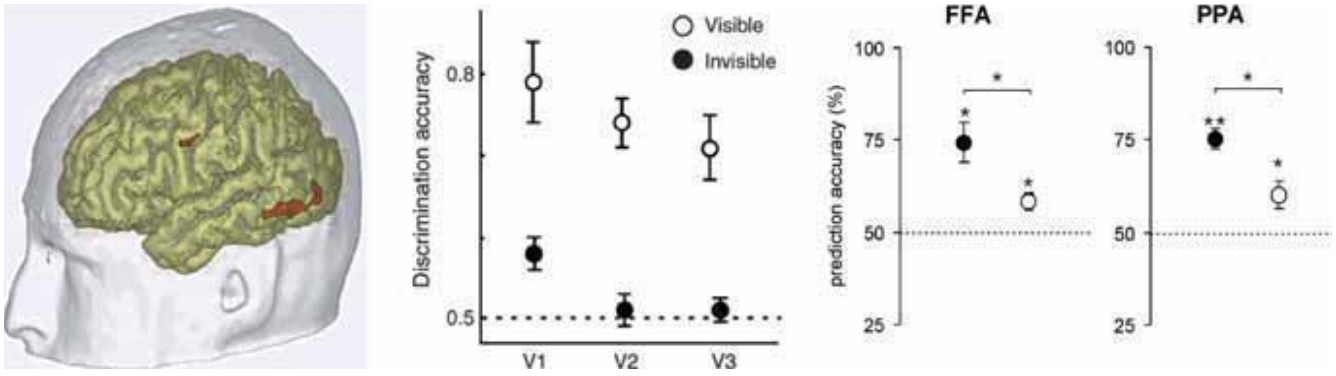


PLATE 97 Activation of sensory cortices by invisible stimuli. *Left panel*: Masked and invisible words nevertheless evoke activation (shown in orange, superimposed on an anatomical image of the brain) of the fusiform gyrus. See Dehaene et al. (2001) for further details. *Middle panel*: Activity measured using BOLD contrast functional MRI in human V1–V3 can be used to discriminate the orientation (right or left tilted) of a grating stimulus. Open symbols representing mean decoding accuracy for a group of subjects (error bars, one SE) for visible stimuli; closed symbols for similarly oriented stimuli rendered invisible by masking. The orientation of these invisible stimuli can still be discriminated at a rate significantly better than chance in human V1. See Haynes and Rees (2005a) for further details. *Right panel:* Performance of support-vector-machine (SVM) classifiers for pairwise classification of face and house presentations from fusiform face area (FFA) and parahippocampal place area (PPA). Average prediction accuracies across participants for visible faces versus houses are denoted by filled circles (±SEM) and for invisible faces versus houses by empty circles. The dotted lines denote chance level (50%). $*p < 0.05$; $**p < 0.005$; n.s. = non significant ($p > 0.1$). Above-chance performance for invisible stimuli indicates that information is still present in higher visual areas sufficient to discriminate stimulus category. See Sterzer, Haynes, and Rees (2008) for further details. (See figure 80.1.)

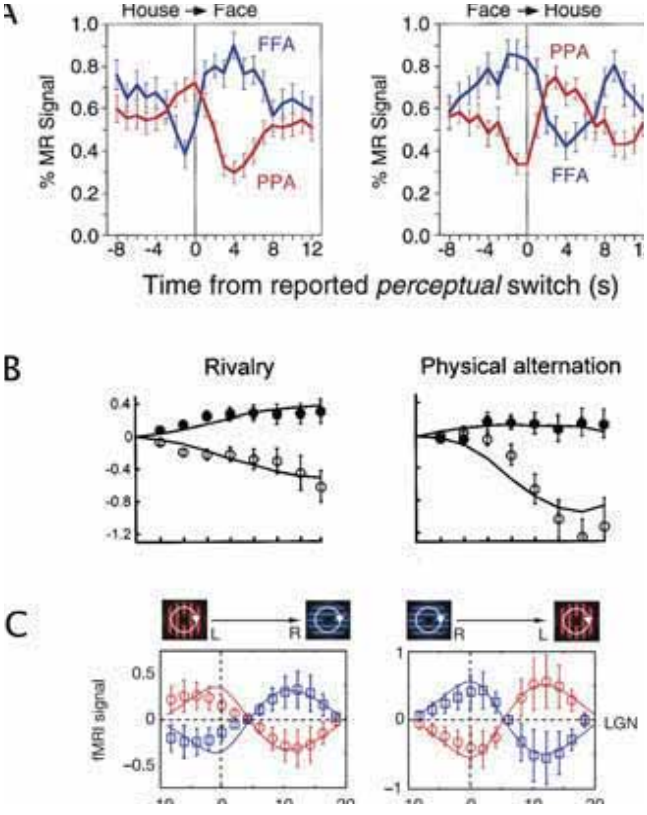


PLATE 98 Fluctuations in activity in visual pathways associated with visual awareness during binocular rivalry. (*A*) Fusiform face area. Activity measured using functional MRI from human fusiform face area (FFA) and parahippocampal place area (PPA) is plotted as a function of time relative to a perceptual switch from house to face (left panel) or face to house (right panel). It is apparent that activity in the FFA is higher when a face is perceived during binocular rivalry than when it is suppressed, and activity in the PPA is similarly higher when a house is perceived than when it is suppressed. For further details see Tong, Nakayama, Vaughan, and Kanwisher (1998). (*B*) Binocular rivalry in primary visual cortex (V1). Activity measured using fMRI from human primary visual cortex is plotted as a function of time after a perceptual switch where the subsequent perception is of a high-contrast stimulus (solid symbols) or low-contrast stimulus (open symbols). The left-hand panel plots activity following a perceptual switch due to binocular rivalry, while the right-hand panel plots activity following a deliberate physical switch of monocular (nonrivalrous) stimuli. V1 activity therefore corresponds to perception during binocular rivalry, and the amplitude changes are similar to those seen during physical alternation of corresponding monocular stimuli. For further details see Polonsky, Blake, Braun, and Heeger, (2000). (*C*) Rivalry in the lateral geniculate nucleus (LGN). Activity measured using fMRI is plotted as a function of time for voxels in the LGN selective for left-eye stimuli (red symbols) or right-eye stimuli (blue symbols) around the time (vertical dotted line) of a perceptual switch between left- and right-eye views (left panel) or right- and left-eye views (right panel). Reciprocal changes in signal in the different eye-selective voxels as a function of perceptual state can be readily seen. For further details see Haynes, Deichmann, and Rees (2005). (See figure 80.2.)

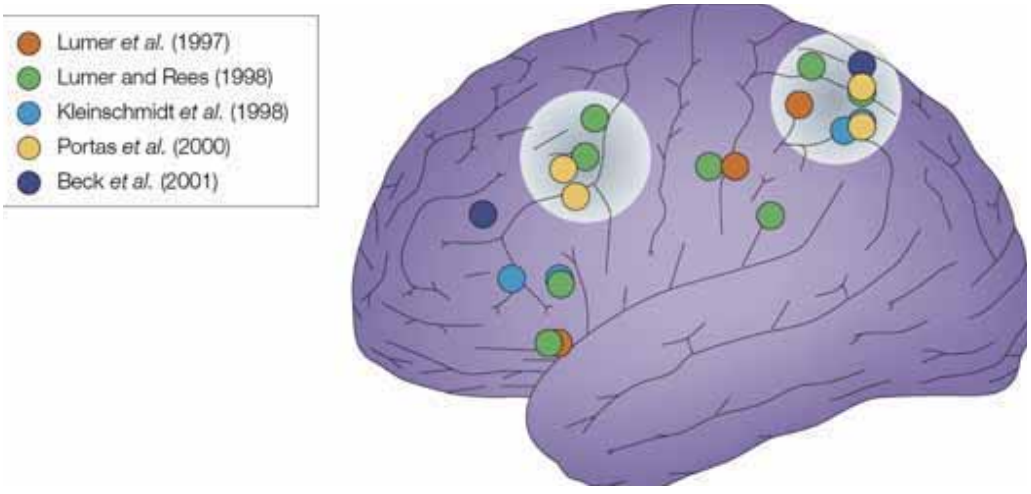


PLATE 99 Parietal and prefrontal correlates of visual awareness. Foci of parietal and prefrontal activity measured using functional MRI and associated with switches in the contents of consciousness independent of changes in physical stimulation are plotted on an anatomical brain image in a standard stereotactic space. Studies shown identify the neural correlates of perceptual switches during rivalry (Lumer, Friston, & Rees, 1998; Lumer & Rees, 1999), during bistable perception generally (Kleinschmidt, Buchel, Zeki, & Frackowiak, 1998), associated with stereo pop-out (Portas, Strange, Friston, Dolan, & Frith, 2000), or associated with change detection (Beck, Rees, Frith, & Lavie, 2001). Clustering of activated foci (white circles) are apparent in superior parietal and dorsolateral prefrontal cortex. (See figure 80.3.)

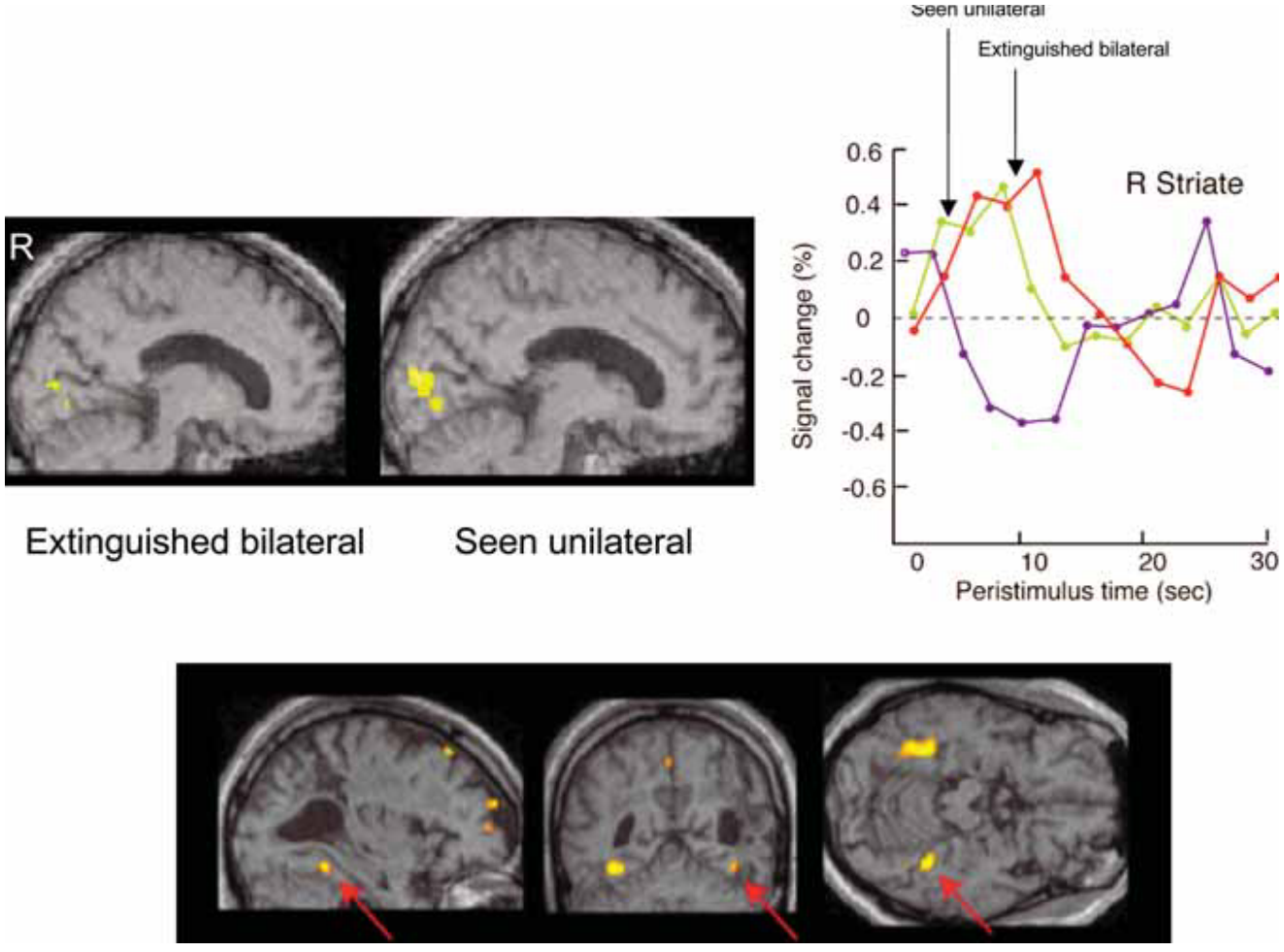


PLATE 100 Activity in ventral visual cortex is not sufficient for awareness. *Upper left*: Activity evoked by an unseen and extinguished left-visual-field stimulus (see text for description of visual extinction) in striate and extrastriate cortex in a patient with parietal neglect. Differences in activity comparing bilateral extinguished with unilateral right stimulation is overlaid on two sagittal slices of a T1-weighted anatomical scan. For further details see Rees and colleagues (2000). *Upper right*: BOLD activity from the right striate (V1) focus of activation shown in the left panel is plotted as a function of peristimulus time for bilateral extinguished stimuli, unilateral left-visual-field stimuli, and unilateral right-visual-field stimuli. Note the similarity of the BOLD time courses for the bilateral extinguished stimuli (in which a stimulus is present in the left visual field but not reported by the individual with parietal extinction) and for the left unilateral stimulation (in which a stimulus is both present in the left visual field and reported). This indicates that stimuli that do and do not reach awareness may produce similar levels of activation in similar cortical locations, in this case following parietal damage. *Lower panel*: Activity in the fusiform face area evoked by a face (versus a house) stimulus presented in the neglected left hemifield of a patient with parietal neglect and left visual extinction. Thus, after parietal damage, activation in the ventral visual pathway for unseen stimuli can be sufficient to distinguish the category of stimulus presented. See Rees and colleagues (2002) for further details. (See figure 80.4.)

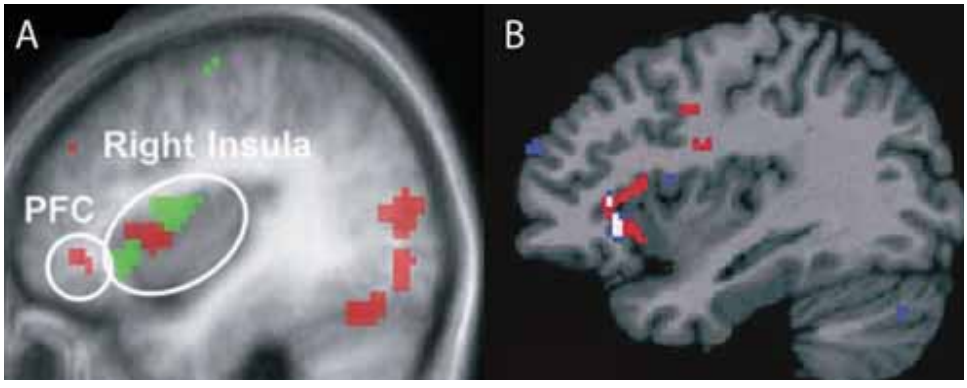


PLATE 101 Functional MRI data supporting simulation theory. (*A*) Brain areas active during the observation of pain in another (red) and the feeling of pain in oneself (green) (Singer et al., 2004). (*B*) Brain areas active during the observation of disgust in another (blue) and the feeling of disgust in oneself (red); overlap in white (Wicker et al., 2003). (See figure 82.3.)